

Psychological Bulletin

SAMPLING IN PSYCHOLOGICAL RESEARCH¹

BY QUINN McNEMAR
Stanford University

INTRODUCTION

One does not have to read much of the current research literature in psychology, particularly in individual and social psychology, to realize that there exists a great deal of confusion in the minds of investigators as to the necessity of obtaining a truly representative sample, describing carefully how the sample was secured, and restricting generalizations to the universe, often ill-defined, from which the sample was drawn. There would seem to be a blind faith in, for instance, the neat formula $\sigma_M = \sigma / \sqrt{N}$, the very simplicity of which belies the fact that certain definite conditions must be met before it is permissible to draw deductions therefrom.

Perhaps the sampling inadequacy of so many researches is merely a reflection of the scanty treatment of sampling in the typical American textbooks on statistical method. Usually, but not always, something is said in the texts concerning the desirability or necessity of securing a representative sample and the possibility of sampling bias, but the specific methods for drawing a sample, checking its representativeness, and avoiding bias are left to the imagination of the reader. This state of affairs is, in part, due to the scarcity of specific techniques for drawing a sample and of methods for checking bias. This scarcity, however, is no excuse for ignoring the fundamental conditions for drawing a random sample, nor does it justify the pro-

¹ This paper was prepared with financial aid from the Social Science Research Council and during the tenure of a visiting fellowship at Princeton University, Autumn, 1938. The writer is indebted to S. S. Wilks for aid on certain mathematical points, but Dr. Wilks should not be held responsible for any errors occurring herein.

mulgation of methods for checking representativeness which are decidedly questionable.

The writer considers it axiomatic that a large amount of psychological research must, of necessity, depend upon sampling for the simple reason that human variation exists. The importance to be attached to sampling will, of course, vary from field to field, and a few investigators may be fortunate enough in their research interests to be able to ignore the problem. It also seems axiomatic that the validity of a scientific inference must depend very largely upon the precision of the data on which it is based. The requisite degree of precision in either the individual measurements or the statistical constants determined from a composite of individuals will likewise vary from field to field. In general, it is desirable to secure the requisite precision in statistical measures with a minimum expenditure of time and effort. The precision of statistical constants, like that of individual measurement, is contingent upon two broad types of errors: random or chance and constant or biased. (The precision of certain statistical constants is also affected by the chance errors in the individual measures, while some, but not all, statistical constants are affected by constant errors of measurement.)

In discussing the problem of sampling, we must keep in mind these two general types of errors, the first of which can be gauged by mathematical formulas, while the magnitude and direction of the second, or biased type, can be evaluated only by thorough acquaintance with, and close scrutiny of, the specific method used in securing the sample. It is the general purpose of this paper to consider available sampling techniques and possible checks on representativeness, and to evaluate the ways by which greater precision in statistical results in either field or experimental work can be attained. More specifically, it is the object of this paper to discuss some of the difficulties of sampling and to consider the applicability in psychological research of the so-called stratified method of sampling. Examples of investigations involving selective factors and investigations typifying adequate sampling will be cited from recent psychological literature. Considerable space will be given to the statistical and sampling aspect of research planning, especially the simple situation involving the use of experimental and control groups. It is not our purpose to discuss sampling as involved in individual measurement, such as time sampling in behavior situations and repeated measures on the same quantity, nor shall we consider the allied problem of sampling of items

for a test or tests for a battery. Neither is it our purpose to include an exposition of the technical mathematics used in the deduction of sampling error formulas.

The general problems and difficulties involved in sampling have been discussed in the texts of Yule (56) and Bowley (1); treated at length in 1926 by a committee of the International Institute of Statistics (16); and more recently discussed from the viewpoint of sociology by Stephan (44), Stouffer (45), McCormick (29), Schoenberg and Parten (39), Woofter (54), and Bowley (3). The modification of the mathematical formulas necessitated by departure from simple random sampling has been treated in the text of Yule (56) and in papers by Bowley (2), Neyman (31), Sukhatme (47), and Wilks (53). The problem of research planning from the viewpoint of statistical method has been given extensive consideration by R. A. Fisher in his *Design of experiments* (9), and Tippett (50) has touched upon various aspects of planning. Psychologists will find that certain parts of the paper by Melton (30) on methodology in learning are devoted to the tie-up between statistical and experimental methods.

GENERAL CONSIDERATIONS

It seems appropriate to discuss briefly certain specific concepts, basic to the general problem of sampling, before discussing the various techniques for drawing a sample and the problem of planning investigations. This section will, therefore, be concerned with the reason for sampling, the nature of the universe being sampled, the concept of homogeneity, experimental hypotheses and permissible statistical inferences, the universality of inductions from samples, control of variation by selection, size of sample, and the fundamental condition of sampling.

Resort is made to sampling because of the difficulty—usually the impossibility—of dealing with an entire universe. The universe is considered as made up of either a finite or an infinite number of units, usually individuals in psychological research. A given investigator may, within limits, define as he pleases the universe which he wishes to consider. Thus, a psychologist may choose the universe of native white 12-year-old boys of urban residence. A sociologist might consider the universe of southern negro tenant families. A universe is said to be finite when there is a limited number of individual units therein and infinite when the number is

unlimited. The standard error formulas for a proportion and for a mean, $\sigma_p = \sqrt{\frac{PQ}{N}}$ and $\sigma_m = \frac{\sigma}{\sqrt{N}}$, assume an infinite universe or population. In the case of sampling from a finite universe these become $\sigma_p = \sqrt{\frac{PQ}{N} \left(1 - \frac{N}{N'}\right)}$ and $\sigma_m = \frac{\sigma}{\sqrt{N}} \sqrt{1 - \frac{N}{N'}}$, where N' is the number of cases in the universe.

In a given research it is sometimes difficult to decide whether the universe being sampled is finite or infinite, and, if finite, it is not always easy to determine the value of N' . It might be argued that psychologists never study an infinite universe. It can readily be seen that the corrective factor in the sampling error formulas becomes negligible as N' becomes large. Thus, if N' is known to be large relative to N , it matters little whether the given universe is wrongly conceived as being infinite. For example, when N is .01 of N' , the term N/N' in the above formulas leads to a reduction in the sampling error of about .005 of its magnitude.

A further distinction between types of universes has been pointed out by McCormick (29). He claims that all sampling in sociology is from a static and finite, or from a dynamic and infinite, universe. Except in so far as one must, when dealing with a dynamic universe, take into account the trends due to its dynamic character, the writer is unable to appreciate McCormick's emphasis on this distinction. The static universe is said to be historical—that is, consisting of past events. It is not noted, however, that an event in a dynamic universe cannot be enumerated until it has occurred, *i.e.* become historical. The determination of the present cost of living is given as an example involving a dynamic universe, while the past cost of living is said to involve a static universe. The only difference between these two situations as regards sampling would seem to be that in making an inference about the past cost of living, the ordinary sampling error formula is applicable, whereas to infer the present cost of living the trend or change must be taken into account. An estimate of *is* requires more information than an estimate of *was*. Another example under *dynamic* is sampling to determine whether divorce is more common among couples when the wife is the youngest child. Since McCormick makes no suggestion as to the use of trends in this problem, the writer finds it a bit meticulous, and practically unnecessary, to classify this problem in divorce as involving sampling from a dynamic, as opposed to a static, universe.

Regardless of the nature of the defined universe, the essential purpose of the sampling method is to provide an economical and feasible scheme for drawing inferences about the defined universe without the necessity of measuring or classifying each individual therein. Here will be found such problems as estimating from a sample the vote in an election, the opinion of a group on some issue, or the frequency of a given type of behavior. In such cases some estimate of the precision of the inference is usually needed; hence, to the various statistical constants standard errors are attached. There is one type of problem involving a sample from a defined universe in which the investigator's chief interest is that of assuring representativeness and in which little concern need be shown for standard errors. We refer to the establishment of behavioral norms for tests or measured characteristics. It is not our purpose to discuss in detail the question of securing adequate samples for norms, since the specific sampling techniques to be described subsequently will be applicable to this situation. There is, nevertheless, one burning issue in regard to norms, especially personality test norms, which we mention in passing, namely: either the failure of the test makers to supply adequate norm information for various groups or the failure of the test users, in their rush to secure psychometric scores, to restrict the use of the tests to individuals belonging to universes for which adequate norms are available. One wonders, for instance, how many psychometric scores for policemen, firemen, truck drivers, *et al.* have been interpreted by the clinician in terms of college sophomore norms.

In psychological research we are more frequently interested in making an inference regarding the likeness or difference of two differently defined universes, such as two racial groups, or an experimental *vs.* a control group. The writer ventures the guess that at least 90% of the research in psychology involves such comparisons. It is not only necessary to consider the problem of sampling in the case of experimental and control groups, but also convenient from the viewpoint of both good experimentation and sound statistics to do so. It is in this connection, as we shall see later, that adequate planning of an investigation yields a statistical as well as an experimental advantage.

The meaning of the term *homogeneous* as used in psychology needs some clarification, particularly as used in describing a sample. Obviously, when a sample is said to be homogeneous, nothing can be inferred from the statement unless it is further stated that it is

homogeneous *with respect to certain* characteristics or variables. Strictly speaking, it is doubtful whether psychologists ever deal with a sample or group which is really homogeneous with respect to a given variable. An exception might be made for such characteristics as age, birth order, and sex, but to speak of homogeneity of a group of men with regard to race, nationality, education, economic status, or cultural background can never imply more than a greater similarity with regard to these characteristics than that found in the generality of all men. Such homogeneity may lead to a very small reduction in the variability of the group with regard to the particular characteristics being studied. For example, to what degree can we expect a great similarity in the spending behavior of men who are homogeneous with respect to incomes, *e.g.* all having incomes of \$3000, unless one also takes into account the size of their families, the nature of their incomes, their place of residence, and other factors which make for a real disparity in effective income?

Another source of confusion, which still exists in certain quarters of the psychological universe, has to do with the type of experimental hypothesis which can be checked by sampling. This, of course, depends upon the kind of inference which it is permissible to make from a sample to a universe. If, for example, one wishes to draw an inference from a sample mean of 60, with a standard error of 1, the only thing that can be said regarding the universe mean is that it is very likely to lie somewhere between 57 and 63, or, if we are willing to be less sure of our inference, we can place the limits at 58 and 62. We cannot specify the probability of the universe mean being between, say, 59 and 61. The more important case, however, for most investigators is that involving the comparison of two means. Let us now discuss this briefly.

The currently accepted rule-of-thumb method is to compute the so-called critical ratio (CR) by dividing the observed difference between the means for experimental and control groups by the standard error of the difference, and then conclude that a nonchance or real difference exists if the CR is greater than 2 or 3. To this we raise no objection, but when a CR of, say, 1.5 is interpreted by saying that the probability of a true difference is .93, or the probable correctness of the difference is .93, or there are 93 chances in 100 of a true difference, we begin to suspect that the investigator has been misled as to the kind of statement which is acceptable in modern statistics. Regardless of the experimental hunch or hypothesis, the only workable statistical hypothesis is that no difference exists

between the universe means. Strange as it may seem, this hypothesis cannot be proven, *i.e.* we can never conclude that no difference exists, but the observable data may force us to reject the hypothesis, and this forms the basis for concluding that *some* difference does exist. This may sound like mere quibbling until one considers more fully the working hypothesis that no difference exists. If the hypothesis is true, then one expects that successive repetitions of the experiment will yield successive differences, the distribution of which will ordinarily be normal with the center at zero and standard deviation corresponding to our observed standard error of the difference. So conceived, we will find by reference to the normal probability table that 14 times in 100, observation will yield CR's of 1.5 or greater. This chance figure, it will be noted, differs somewhat from the 7 times in 100 which might be inferred from the .93 above. On the basis of the given hypothesis, we can make a rigorous statement as to the probability of obtaining a difference as large as our observed difference, but on no conceivable hypothesis can we make a probability statement concerning the true difference. For a thorough discussion of this point the reader is referred to Fisher (9), who has introduced a useful concept to which we now turn.

The concept of "fiduciary limits" of Fisher, and its equivalent "confidence limits" of Tippett (50), permits one to infer from a sample mean (M) that the universe mean is between M plus and minus either $2\sigma_M$ or $3\sigma_M$ or some other multiple of σ_M . The correct multiple is arbitrary, but it should be noted that the degree of confidence varies with the multiple we use. The same reasoning applies to inferences regarding a population proportion, the difference between proportions, the difference between means, etc. In the case of the difference between statistical constants (say, means), one takes $D_M \pm 3\sigma_D$ as limits, and from these limits concludes not only that a real difference exists (if the lower limit is greater than zero), but also something as to the likely magnitude of the true difference. In our eagerness to conclude that a real difference exists, we too frequently ignore the important fact that something can be said concerning its magnitude. The reader will have noticed that for a given degree of confidence, the fiduciary limits can be narrowed only by decreasing the size of the standard error. This greater precision can be secured either by increasing the size of the sample or samples or by alterations in the methods of drawing the sample. This will receive detailed discussion later.

It is also necessary to keep in mind that a sample from a defined

universe permits an inference about that universe and no other. One cannot generalize beyond the universe from which the sample was drawn unless it can be demonstrated that the given universe, and therefore the sample from it, is typical of some other, perhaps more general, universe. The extent to which a universe is limited, *i.e.* does not include the generality of all human beings, involves the notion that it is relatively homogeneous in certain respects. One might readily grant that much can be gained by limiting the universe in such a way as partially to hold constant certain variables, but there is a limit to this type of procedure. We do not find ourselves in agreement with Peatman's (33) recent argument favoring the selection of samples that tend to be homogeneous in certain characteristics and thereby limiting our generalization to populations also homogeneous in the chosen characteristics. Peatman goes on to say: "It is possible by the method of homogeneous limitations to establish samples of subjects which will tend to be more fair for a given psychological problem than if the method is not used." Since a limited generalization is limited, and since so many of the generalizations of psychology have been, and are, circumscribed because of the restricted nature of the investigated universe, we are inclined to suggest less, rather than more, restraint as regards the universes defined for study. It may very well be that a structuralistic psychology could draw valid and sweeping generalizations from research on a few highly selected, highly trained individuals, but it is difficult to see the value of generalizations based on college sophomore samples when the enquiry is concerned with the typical topics in social psychology and the psychology of learning and of individual differences. Whether the amazing array of information accumulated about the college sophomore, regardless of its possible value to psychologists and others as pedagogues, is of any great value for describing, predicting, or controlling the generality of human behavior is a debatable question.

Aside from the necessary restriction in generalizations which results from the use of limited universes, there is also the danger that the selection of subjects by the so-called method of homogeneous limitation may distort research results, especially in studies involving the correlational method. An example of a vitiating type of selection is to be found in a study of assortative mating (38), which is based on 46 couples claimed to be "strictly homogeneous." It is said that "in insisting on strict homogeneity three results have been achieved: the disturbing effect of the presence of extremes on the correlation coefficients have [sic] been avoided, the group used is

very representative of its particular segment of the whole population, and the correlations obtained for the control group are more significant." We are at present concerned only with the first of these three results, but incidentally it should be noted that no evidence is given to support the second claim and that it can be said with regard to the third that correlations based on a control group formed by pairings at random can never have more than purely chance significance and therefore possess only pedantic value. As to the first claim, it should be noted that the selection of a group relatively homogeneous as regards age, education, occupation, socioeconomic status, and religion automatically reduces such assortative mating coefficients as may exist for traits which are related to these characteristics for which homogeneity is claimed. When studying trait variation and covariation, care must be exercised in homogeneous selection with respect to variables other than age, sex, race, and nationality lest we unduly disturb the variation and covariation of the very traits being investigated. Holding variables constant experimentally may involve one of the fallacies in the use of the partial correlation technique, *i.e.* it is possible in some cases to hold too much constant.

The fundamental condition for random sampling is that each unit or individual of the defined universe must have an equal chance of being drawn, and, once drawn, no unit can be discarded without risk of bias. In psychological research, individuals are apt to be discarded because of incomplete information, or an individual may discard himself by refusing to coöperate. Because of the extreme difficulty of assuring that each individual or unit has an equal chance of being included in the sample, Bowley (3), an English statistician, has expressed extreme skepticism of sampling and the use of sampling error formulas. Any failure of this condition for simple (sometimes called Bernoullian) sampling will lead to bias and therefore to a biased inference regarding the universe from which the sample has been drawn; or, when two universes are being compared, the presence of bias in one sample or both may lead to an obtained difference which, rather than being real, is actually due to selective factors.

A requisite for the use of sampling error formulas when variables, rather than attributes, are being studied is that the distribution of scores or individual measures shall be approximately normal or at least not too markedly skewed. Just how much skewness is permissible seems open to debate; psychologists dealing with variables yielding skewed distributions are in need of an expository paper on this problem.

Another persistent question, perhaps deserving a short paper, has to do with the size of the sample. How many cases should one use? Obviously, there can be no one set answer to this question, not even the time-worn advice to secure as many as possible. The number of cases required must be based upon the desired degree of precision or permissible magnitude of error, which, in turn, is dependent upon the nature of a particular investigation. If the task is to indicate the presence of some attribute in a group with a given margin of error, one can readily ascertain the number required for the given degree of precision. If two groups are being compared on some variable, the sample size may be determined by an intuitive hunch as to the possible magnitude of the difference between the two universes. One rule which can be followed with comparative safety is that the demonstration of a difference (or effect) which is large enough to possess any practical or social significance will not require large samples; certainly, a difference which is so small as to require 1000 cases in each sample to demonstrate it is apt to possess little psychological meaning. Researchers who attempt to show that correlation exists or that two correlation coefficients are statistically different will usually need a rather large number of cases to establish positive results.

Some psychologists frown upon the use of small samples, as, for example, N less than 25; a few use such small samples, but scorn the necessity of evaluating their results in terms of the mathematics of small samples ("the very idea of using statistical refinement with so few cases . . ."); while others will rightfully argue that when small samples, properly evaluated, yield a difference which would arise by chance only once in a hundred times, the result is just as dependable as if the same chance figure had been found for large samples. It is assumed in either case—small or large sampling—that the sampling technique is such as to avoid bias. It is commonly and erroneously thought that some magic lies in large samples and that bias is less apt to be present. The larger the sample, the greater the precision so far as random errors are concerned, but it does not follow that bias is avoided by increasing the size of the sample.

SAMPLING TECHNIQUES

In considering the specific methods of drawing a sample so as to avoid bias, we must differentiate between two types of situations: (1) All the units or individual members of a given universe may already be catalogued or on file with more or less information of

some kind already known concerning the universe; or (2) no file is available, and little is known about the universe except what has been inferred from previous samples. The first is typified by the universe of telephone subscribers, or those on relief rolls, or the school population of a city, while the second is the typical universe dealt with in field surveys and investigations, such as the straw and public opinion polls.

Sampling methods, as used, may be classified under four headings: accidental, random, purposive, and stratified. These will be discussed in the above order with more attention given to the second and fourth methods.

Despite the fact that psychologists seem to use the *method of accidental* sampling more than any other, it has nothing to recommend it either on statistical or scientific grounds. Its very ease and simplicity have, no doubt, led to its wide use. This method is essentially nothing more than its name implies: the accidental choice of individuals for the sample. Any individual who is available and can be corralled into service becomes a subject. The method has its corollary in the haphazard and accidental manner in which many universes are chosen for study. In fact, the available subjects may not have been chosen as representing any defined universe, but used to define *a posteriori* the universe being sampled. It is here that the college sophomore has an advantage in being the raw stuff out of which psychologists build a science of human behavior. Aside from the failure of the characteristics of sophomores to be typical of the generality of mankind, one must also remember that the lowly soph is of a decidedly different species as we pass from institution to institution. Even granting that the college sophomore is typical of mankind, certain accidental factors affect the likelihood of any one individual's inclusion in a sample of sophomores. His coöperation must be secured, and, what may be more important in personality studies, his chance of representing *Homo sapiens* is increased if his interest in himself and his own personality adjustment has led him to take elementary psychology.

Accidental sampling also takes place in more serious attempts to secure a fair sampling of some defined universe. Public opinion polls and all questionnaire studies which depend upon the voluntary coöperation of people will be affected by accidental sampling. That some of the factors operative for questionnaire reply are highly pertinent, though accidental in nature so far as the unwary investigator is concerned, is brought out by Crossley (6), and by Katz and

Cantril (21) in their discussion of the straw polls of 1936. It should be noted that these accidental factors are not necessarily purely chance in that they may operate differentially so as to lead to the exclusion or inclusion of particular individuals.

By the *method of random* sampling it is fairly easy to arrive at a representative sample, provided the universe has already been catalogued. Thus, if one wishes a sample of school children of a certain grade in a city, one can secure a representative sample by a purely mechanical scheme, such as taking every *n*th card from the files. This will assure a random sample unless the cards have been systematically arranged in other than alphabetical order.

A psychologist will find little consolation in the thought that there are mechanical schemes for drawing a random sample, since files seldom exist for the universes with which he deals. The use of the random method for sampling an uncatalogued population involves so many difficulties in psychological research that no specific schemes are to be found in the literature. That the hand-picking of units at random by eye may lead to bias in the relatively simple problem of selecting wheat shoots has been pointed out by Yates (55). The personal selection of cases in psychological work may also lead to bias, as, for example, the selection of preschool children in the New Revision of the Binet (48), which was so obviously biased that the records for a large number of cases had to be discarded. The *Literary Digest* straw polls rested on the assumption that the population of telephone and car owners were not different in their voting preference from the entire population of potential voters. This happened to hold prior to 1936, so that replies to ballots mailed at random to telephone and car owners forecasted fairly accurately the election results. The failure in 1936 is attributed to a change in the alignment of voting to class or income lines.

Because of the difficulty of devising a scheme which permits each individual of an uncatalogued universe an equal chance of being included in the sample, investigators have resorted to purposive and stratified sampling in their efforts to secure fair and unbiased samples. Many psychologists have used something akin to stratified sampling, but nowhere in the research literature of psychology does one find any hint that such methods disrupt the fundamental condition of simple sampling and that consequently the ordinary sampling error formulas are in need of modification.

The *purposive method*, as the writer understands the rather inconsistent statements thereof, depends upon the selection of groups which,

together, yield the same averages or proportions as the whole universe with respect to those quantities or qualities which are already a matter of knowledge. If the variables under study are related to the known factors, the samples (groups taken together) will be typical of the whole. It should be noted that all the individuals in the several groups are used, that the sampling unit is the group, that the efficacy of the method depends upon the degree of relationship between the criterion variables and the characteristic being studied, and therefore that its use is contingent upon considerable foreknowledge. The method is essentially one of weighted averages, and according to Neyman (31) it is not very reliable. Since the method has not found much favor and since it is not particularly adaptable for psychological sampling, we will give it no further consideration. The interested reader can turn to the discussion of Jensen (16) and the more technical paper of Bowley (2).

In the *stratified method*, one or more individuals are pulled at random from each of several strata, the number in the sample from each stratum being proportional to the universe number in the stratum, and the strata are predetermined by known knowledge on some control variable or variables. Psychologists who sample so as to secure proportionate representation from the several occupational levels are, in reality, using the stratified method. It should be obvious that the method can be used for either catalogued or uncatalogued universes, providing information is available on some variable or variables which permits their use in setting up the strata. Common-sense reasoning and mathematical treatment agree in showing that the method gives more reliable results than the purely random method, providing the experimental variable is related to the stratifying variables. Thus, if we had information on some universe with regard to the heights of the individuals, nothing would be gained by using height as a means of setting up strata for the purpose of drawing a sample from which to infer the IQ's of the group. Such a procedure would not lead to better (or worse) results than would be obtained by the random method.

There are three reasons why it is convenient at this point to present the formulas for the sampling errors involved in stratified sampling. A consideration of the formulas will indicate (1) that they are different from the ordinary formulas, (2) that greater precision results from stratified sampling, and (3) that there are limiting factors as to the possible increase in precision. It might be anticipated that the error formulas for stratified sampling would differ

from the ordinary formulas in that the condition of sampling is essentially different. The formulas themselves indicate greater precision for the stratified method, and it seems reasonable to assume that a sample drawn by the method would be less subject to bias, since by it one tends to have all strata, or groups, or levels, represented in the proper proportions.

The formulas which follow have been culled from the papers on the mathematics of stratified sampling. We are not giving the necessary variations for sampling from finite universes for the simple reason that there is scarcely any practical advantage in these forms over the close approximations yielded by those which assume an infinite universe.

When sampling for attributes by the stratified method, the standard error of an obtained proportion, P , is given by

$$(A) \quad \sigma_P = \sqrt{\frac{PQ}{N} - \frac{\sigma_p^2}{N}}$$

where P equals the proportion in the total sample, N , who possess the attribute, $Q = 1 - P$, and σ_p is the weighted standard deviation of the several strata proportions about the sample value, P , or

$$\sigma_p^2 = \frac{1}{N} \left[N_1(p_1 - P)^2 + N_2(p_2 - P)^2 + \dots + N_k(p_k - P)^2 \right]$$

where N_1 , N_2 , etc. are the number of cases, and p_1 , p_2 , etc. the proportions, in the several strata, there being k -strata in all. A casual examination of (A) indicates that the magnitude of the error for a stratified sample is less than for ordinary sampling, and that the increase in precision depends upon one's ability to stratify the universe in such a way as to secure strata which are really different with regard to the attribute being studied. For example, if voting did not follow class lines, nothing would be gained statistically by sampling separately the several socioeconomic levels; or, if the vote tended to be the same in all states, it would be unnecessary to sample each state separately. In this case, random sampling of any one socioeconomic level or of any one state will yield just as accurate results as stratification.

The formula for the standard error of the mean when the sample has been secured by the stratified method has been stated variously by Yule (56), Bowley (2), Neyman (31), Sukhatme (47), Woofter (54), and Wilks (53). We give herewith a few variations in simplified notation. Any reader who prefers more elegant expressions can refer

to either Neyman or Sukhatme. The variance (standard error squared) of the mean is given by

$$(B) \quad \sigma_{\bar{x}}^2 = \frac{1}{N} (\sigma^2 - \sigma_{\bar{x}_i}^2)$$

where \bar{x} = the sample mean, σ^2 = sample variance, and $\sigma_{\bar{x}_i}^2$ = the variance of the means of the several strata about the total sample mean. An exactly equivalent form is

$$(C) \quad \sigma_{\bar{x}}^2 = \frac{1}{N} \left\{ \sigma^2 - \frac{1}{N} \left[N_1(\bar{x}_1 - \bar{x})^2 + N_2(\bar{x}_2 - \bar{x})^2 + \dots + N_k(\bar{x}_k - \bar{x})^2 \right] \right\}$$

where N_1, N_2, \dots are the numbers and $\bar{x}_1, \bar{x}_2, \dots$ the means in the separate strata. Expression (C) states explicitly that the term $\sigma_{\bar{x}_i}^2$ of (B) involves weighting each stratum mean by the sample number of cases in the stratum.

If stratification has been accomplished by the use of a characteristic or variable, u , which is linearly related to the variable being studied, the formula can be written in the form

$$(D) \quad \sigma_{\bar{x}}^2 = \frac{1}{N} (\sigma_x^2 - \sigma_x^2 r_{xu}^2)$$

or, if one prefers, he can compute the standard deviations separately for the several strata distributions of the variable being studied and use these to arrive at the standard error of the mean by substituting in

$$(E) \quad \sigma_{\bar{x}}^2 = \frac{1}{N^2} \sum N_i \sigma_i^2$$

or its equivalent

$$(F) \quad \sigma_{\bar{x}}^2 = \frac{1}{N^2} (N_1 \sigma_1^2 + N_2 \sigma_2^2 + \dots + N_k \sigma_k^2).$$

It matters little which of these formulas is used in practice, except that (D) is not so general as the others. It can, however, be made so by substituting the proper "eta" for r . Perhaps form (B) is the more practicable. Regardless of the form, it will be noticed that stratified sampling does lead to greater precision in the sense of a smaller chance error, but this is only so when the control or stratifying variable is related to the variable being studied. This is explicit in (D) and directly implied in formulas (B) and (C), *i.e.* the means differ from stratum to stratum, and form (E) indicates the increase in precision, if any, as due to greater homogeneity for the variable being studied within the several strata than that which exists for the total sample. These are but three slightly different ways of regarding the same thing.

One can also deduce from the above formulas that stratification on the basis of a variable, u , for studying variable x may not lead to an improved sample for studying some other variable, y , unless x and y are correlated. If several variables are used in stratifying, the correct standard error formula involves substituting in (D) in the place of r_{xu} the multiple correlation between x and the control variables. Since the multiple correlation coefficient increases slowly as more variables are added, it follows that the gain in precision which results from using more than two or three control variables may be very small.

The applicability of the stratified method depends, of course, upon *a priori* knowledge of the universe with regard to possible control characteristics, and its advantage is contingent upon the additional condition that the variable being investigated is related to the possible control variables. Often, information is lacking on this latter point, so the investigator must rely on judgment as to what variable or variables will make profitable controls. At the present time the characteristics which can be utilized as controls in stratified sampling in psychology are few in number: socioeconomic or occupational status, urban or rural, geographical factors, age, sex, racial or national origin, and perhaps intelligence, and education. Strata can be established upon these with approximate knowledge concerning the proportion of the entire population falling in the several strata, but these proportions may not hold for restricted, and more commonly used, universes. Despite the limitation of the stratified method of sampling, its use offers psychologists the best available scheme for drawing a representative sample. In addition to yielding a possibly greater precision, the method should, perhaps, tend to the elimination of bias. Examples of its use in psychological research will be cited later.

CHECKING REPRESENTATIVENESS

Once a sample has been drawn, particularly in field investigations where mechanical schemes cannot be utilized, the investigator may wonder whether it is really representative of the universe from which it was drawn. At least three recent statistical texts offer methods for checking representativeness. Sorenson (42, pp. 320-321) suggests two methods: See whether adding additional cases changes the value of the statistical measures, and draw additional samples and compare results with those obtained from the original sample; while Smith (40, p. 317) says that "the only test of adequacy [representativeness] of

the sample (in the absence of *a priori* knowledge by which homogeneous classification could be made) is to take several random samples and see whether or not the results approximate the same each time." Garrett (11, p. 243) also suggests this latter method. Just how either of these schemes can be expected to yield an answer to the question at issue is a bit mystical. Suppose one finds that a second sample does give results in agreement with the first sample; what does it prove beyond the fact that samples drawn in the same manner will agree within chance limits? Any concealed bias in the sampling method will never be detected by such a procedure. The first method proposed by Sorenson will also fail, since additional cases drawn by the same method will be subject to the same bias as the original sample if bias were present therein.

Splitting the sample into halves and comparing means for the two halves is a slight variation of the above-mentioned methods. It should be obvious that bias will not be detected by this scheme—each half is affected by the same factors as the whole. Closely allied to these schemes for checking representativeness is a method advocated for determining whether the size of a sample is adequate. If two random halves yield means which are not significantly different statistically, the sample is said to be adequate in size. The trouble with such a criterion is that the two halves of a sample of 100 will, in the long run, yield CR's of the same magnitude as will be obtained by comparing the two halves of a sample of 1,000,000.

One might, at this point, raise a question as to the value from the experimental viewpoint of repeating one's own investigation. When replication discloses sources of error, no one would deny that repetition was worth while. In case the repetition of an experiment leads to a duplication of the results originally obtained, the investigator is apt to place much more reliance on his findings. Now it might be gratifying to an investigator to know that he can duplicate his results, but since any experimental (or sampling) bias may also be duplicated, the ultimate value of such confirmation is problematical. The recent work on telepathy and clairvoyance illustrates the point that mere iteration of one's own research is not sufficient in the establishing of acceptable scientific facts. Psychologists have been too slow in repeating the work of their contemporaries, and not a little of this dearth springs from a blind faith in probable errors.

The only adequate method for checking the representativeness of a sample is to compare its results with the universe values, but, of course, it is very seldom that the universe values are known.

Hilton (15) discusses a case in which a sample of 1% of a million yielded values in close accord with those found for a 33% sample drawn by a different method from the same finite universe. It is possible for the conductors of straw polls to get a fairly satisfactory check on their techniques by comparing their final predictions with the actual votes. Found discrepancies may, however, be due to changes in voting preferences which occur in the interval between the straw poll and election day. The success of certain of the straw polls, however, tends to confirm one's faith in sampling.

In the absence of any rule-of-thumb method for checking representativeness in psychological research, the investigator must resort to logical considerations. If the sample has been drawn by some mechanical means or by stratifying the universe on the basis of pertinent facts, one can feel fairly sure that the sample is representative. In the absence of an obviously valid scheme for drawing the sample, the only thing one can do is to describe the sample as completely as possible with regard to known characteristics of the universe from which it was drawn. If the sample is typical of the universe in several variables which are related to the variate being studied, it is safe to assume that it is representative. This reasoning is, of course, posterior use of the principles of stratified sampling. The importance of fully describing the sample and how it was drawn cannot be overemphasized. Without such information it is impossible to evaluate a given research.

EXAMPLES OF ADEQUATE SAMPLING

In the absence of any very specific positive suggestions as to how a representative sample can be drawn, it will be instructive to consider examples from the literature which, in the opinion of the present writer, exemplify good sampling procedures. The interested reader will wish to turn to the references cited in order to obtain more detail than it is feasible to give here. As an example of an extensive sampling project we may refer to the report by Schoenberg and Parten (39) in which are outlined the sampling methods and difficulties involved in the Urban Study of Consumer Purchases. The sampling unit for the study is the family. The first step was to draw a random sample by the use of city directories, etc. of 625,000 families. Schedules prepared for these cases contained information on nativity, color, family composition, and housekeeping arrangements; for a subsample, selected randomly, of approximately 250,000 the following information was obtained: income, occupation, composition of family,

type of living quarters, home tenure, and rentals. This group of 250,000 met certain eligibility requirements as to nativity, color, and family composition. Then, to the subsample were added families selected on the basis of stratification so as to secure a better representation of the salaried and independent professional and business groups. The final sample of 30,000 was then chosen by the stratified method, the stratification being on the basis of the information obtained from the 250,000. From these 30,000 cases detailed information on expenditures will be secured.

Few, if any, psychologists will ever be in a position to follow such an elaborate sampling scheme. The study of Garrett, Bryan, and Perl (12) indicates one effective way of sidestepping the sampling problem. They used *all* the 9-, 12-, and 15-year-olds in a small city school system in a study which depended upon comparison of these three age groups in regard to the intercorrelations of tests. It cannot be argued that the found differences are due to selection. Another example of taking an entire school population as the sample is to be found in the study of intelligence and birth order by Steckel (43). It should be noted, of course, that the nature of these studies is such that one would expect small variation in the results as one passes from city to city, and rather than attempt to draw individuals for representative samples, these two studies depend more upon the choice of typical cities.

A similar scheme was followed by Jones and Conrad (20) in their study of the growth and decline of intelligence. Their monograph not only illustrates adequate methods for surmounting the difficulties of sampling, but also includes a thorough description of the group used and an enlightening exposition of possible selective factors in studies of this type. This study will repay careful reading by those who are interested in the sampling problem.

Perhaps no field of psychology is fraught with such complex sampling difficulties as those found in studies of race and nationality differences. The critic and skeptic can point to some selective factors in nearly every study in this field. That creditable research can be done in this field is demonstrated in the excellent studies by Klineberg (23) and by Franzblau (10). Both investigations are good examples of adequacy as regards their treatment of sampling so as to avoid bias or selection, and both give ample information as to how the groups were chosen. In fact, the writer knows of no other such serious efforts to avoid bias in this field of research.

Those who are interested in sampling for the purpose of estab-

lishing test or behavior norms will find the recent work of Terman and Merrill (48) a good example of extensive sampling so as to secure age samples fair for the generality of children in the United States. Stratified sampling on the basis of geography, urban or rural, socioeconomic status, plus localities judged typical of particular sections of the country, tends to make their age groups more representative than any ever before obtained. The greatest difficulty was encountered in securing fair samples for the preschool groups and for the upper ages.

Since the results of the Gallup poll and the *Fortune Magazine* poll of 1936 checked fairly well with the final election outcome, we will here indicate briefly the sampling schemes they used. More detail can be found in the papers of Katz and Cantril (21) and Crossley (6). Gallup's procedure involved a carefully chosen mailing list which was supplemented by interviewing other individuals (one-third by interview; at present the Gallup polls are entirely by interview). The factors considered were a state's population, ratio of farm to city population in each state, income levels, age, correct proportion as to those who voted for Roosevelt, Hoover, and Thomas in 1932. The interview procedure was used to supplement the number secured by mail from the lower income brackets and to counteract the following factors which apparently operate in the return of mailed ballots: People with intense opinion (reformers, arch-conservatives, radicals) are more apt to reply; educated people take greater interest; the economically secure feel more free to reply; and men are more apt to reply than women. It is thus seen that Gallup is using the stratified method, and his greater accuracy in predicting 1938 election results is evidence, perhaps, that improvements have been recently made in his methods. The *Fortune Magazine* poll, which outdid its rivals in 1936, depended upon interviews by an unusually able staff who sought out a few (relatively) typical voters who presumably were characteristic of a large group to which they belonged. Whether this method of sampling is actually adequate, or the 1936 results just lucky, will not be known until its further use.

An interesting procedure for circumventing a part of the difficulties due to selective factors is to be found in a paper by the late E. A. Robinson (35). A sample of 8419 voters was admittedly not representative, but when broken down into party, sex, and occupational groups, enlightening comparisons were made by separate treatment of the possible subgroups. Thus, the Republican *vs.* Democrat *vs.* Socialist voter attitude, *e.g.* toward "currency stability," was

determined for subgroups according to sex and a sixfold classification of occupational status. The advantage of such a procedure is that it avoids bias or selective factors due to the named groups and hence is better than comparing, say, Republicans and Democrats as total groups. This procedure also avoids the possible masking of important facts, which all too frequently occurs when large heterogeneous groups are compared. There are two limitations here: The sampling for each subgroup must itself be random, and the grand total must be sufficiently large to avoid too few cases in a subgroup.

Other studies could be cited in which the problem of sampling has received adequate attention, but the writer was disappointed in the fewness of such studies in the literature of social (broadly defined) and educational psychology for the past 10 years. It should be remembered, however, that we have so far been concerned only with sampling in those field investigations in which an inference is made from a sample to a universe and those field studies involving the comparison of different groups, and that the literature does not abound in researches of these particular types.

SOME EXAMPLES OF SELECTIVE FACTORS

One way of avoiding bias due to selective factors is to profit by the difficulties and errors of others. We will, therefore, in this section give a few examples of recent studies in which selective factors have operated or in which the problems associated with sampling have received inadequate handling.

Woofter (54) has cited some instances in which results based on a *total* universe have been needlessly interpreted *via* probable errors, but when he claims that a study (reference not given) of whites and negroes should not have been evaluated in terms of sampling because the subjects chosen for study and testing included *all* the 12-year-olds in three white and two colored schools of Nashville, we are inclined to feel that he is laboring under a misapprehension. Woofter argues that the study permits a complete induction and that any differences so discovered must be significant, but that the differences apply only to the finite universes under consideration. If his argument is valid, one would not be permitted to think of the next year's crop of 12-year-olds in these schools, nor would one be able ever to consider particular groups as typical of the larger universes to which they belong.

An example of a selective factor, involved in a sampling study of birth rate, has been given by Kiser (22). Fieldworkers failed to

revisit families missed because no one was home. This led to a bias, because the wife who is away from home is apt to be childless or have fewer children than the average. Evening revisits demonstrated the bias. In a recent study (34) of ordinal position as related to psychological traits, one finds criticism of others for using atypical samples and an argument that a "normal" sample is necessary. This investigation was based on two groups of sixth-graders with average IQ's of 112 to 114. Are these not also atypical? And is it any safer to draw an induction from these groups to the generality of "normal" children than in some of the cases he criticizes? Incidentally, it is stated in the conclusion that the "representativeness of each of the two samples has already been demonstrated." We have looked in vain for evidence justifying this statement.

The danger inherent in voluntary coöperation is illustrated in a factor analysis study (27) of Spranger's value-types. Of 600 papers distributed to students, only 265 were returned. It is said that "the selective factors at work in determining which students answered and which discarded the tests were of no importance in this study, as it is only the interrelations of the scores on the various items with which the investigation was concerned" (p. 19). What of the likely possibility that those who were not of the "theoretical," cognitive, rational, scientific type tended to discard the test papers? This, as a selective factor, would affect the homogeneity of the group with respect to this characteristic, and hence the intercorrelations; and therefore self-exclusion from the sample might have introduced a selective factor which cannot be so readily dismissed.

It is well known that the research on the problem of later maturity is complicated by a tangle of selective factors in that it is difficult to secure fair samples for the later decades of life. A recent study (41) illustrates how an unnoticed selective factor can creep in. It was found that total aptitude scores increase with age up to 60 years, that vocabulary scores also increase, and that paragraph reading shows no change with age. Since these results are inconsistent with previous findings, the investigator proposes an explanation in terms of *use*, since his group consisted of individuals, teachers, and part-time students who had continued using their intellects. Cannot it be said that individuals who so continue represent a selection? More important, however, is the fact that in equating the age groups for occupation and amount of schooling a definite selective factor favoring the older is introduced. This is true because, in general, the amount of schooling of adults varies with age, and those of a gen-

eration ago who received a college degree were very likely superior to more recent run-of-the-mill college graduates.

SAMPLING AND THE USE OF EXPERIMENTAL AND CONTROL GROUPS

The problem of securing precision and avoiding bias due to sampling offers somewhat different possibilities in studies of an experimental nature than in those field investigations which involve the comparison of two groups. In the latter situation one can draw random samples from the two universes and depend upon large numbers for the reduction of sampling errors, or one can run less risk of bias and secure greater accuracy by drawing the samples by the stratified method. In the experimental situation, one can depend upon randomization as a method of balancing out chance factors and upon large numbers as a means of reducing the magnitude of the standard error of the difference (between means, say), or one can avoid bias by holding constant relevant factors by the method of pairing individuals and at the same time reduce the sampling error without necessarily using larger numbers in the groups. Just as in the case of stratified sampling, where an adjustment in the standard error formulas must be made because we are no longer dealing with simple sampling, so in the case of building up an experimental and control group by the use of pairs related on relevant variables we must also make some allowance for the fact that we have interfered with the principles of simple sampling. The correlational term in the standard error of the difference formula

$$(G) \quad \sigma_{D}^2 = \sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2 - 2r_{12}\sigma_{\bar{x}_1}\sigma_{\bar{x}_2}$$

is necessary and sufficient for making due allowance for the fact that the two samples have not been drawn independently of each other.

It is also well known that the correlational term, being subjective, represents the statistical advantage which accompanies the experimental advantage of control by the use of pairs. Such control should, of course, lead to greater precision. The most frequent statistical error in the psychological literature is the failure to use the correlational term in the above formula when the situation demands its use.

In the planning of research wherein an experimental and a control group are essential, it is well to consider rather carefully the benefits to be derived from control of variables, likely to be sources of error, by way of pairing or matching with respect to these variables *vs.* depending upon randomization as a method of controlling these

factors. The pairing of individuals as a method of equating an experimental and control group with respect to relevant variables which might be related to the experimental variable has long been recognized as sound experimental method. Any found difference between the two samples cannot be explained as due to a difference between the groups in regard to the variables so controlled. The ideal experimental situation would be attained when all the variables likely to affect the difference between the groups on the experimental variable were controlled in the sense of being equated for the two groups. But so little is known about the interdependence of psychological variables that this ideal can never be achieved. It follows, therefore, that regardless of how carefully we equate two groups on the basis of certain variables, there will be other variables of more or less importance upon which the groups might differ, and the only hope is that by the principle of randomization no greater than chance differences between the experimental and control groups will exist for these unknown variables.

Consequently, one may very well ask whether there is an advantage in equating by pairing. The answer is "yes," providing one's knowledge and intuition are such that out of the available variables for pairing one can select those which are really pertinent. If one is fortunate enough in pairing to create an interpair correlation on the experimental variable as high as .75, the standard error of the difference between means will be reduced by one-half. To accomplish this increase in precision by using larger groups would involve quadrupling the original numbers. An r of .50 will increase the precision as much as will doubling N . On the other hand, if equating does not lead to pair correlation on the experimental variable, one has evidence that the pairing scheme, regardless of how elaborate, has yielded neither a statistical nor an experimental advantage. There remains only the psychological satisfaction of knowing that certain variables were controlled.

If the experimenter has no hunch as to what variables should form the basis for pairing, he must depend solely upon the principle of randomization, which, in the typical situation, consists of dividing a given group randomly into halves and taking one half for the experimental, the other for the control group. If the available group of individuals can be catalogued in some fashion, it can always be split into halves by some mechanical scheme, thus assuring randomness with regard to *all* the known or unknown characteristics of the individuals. If the experimental cost per individual is such that

fairly large numbers can be utilized, this scheme of randomly splitting a group into two subgroups as experimentals and controls has much to recommend it. The experimentalist may object to this by saying that he prefers not to trust chance or luck to yield two groups which are comparable on what he thinks are pertinent variables. In this connection it is important to remember that randomization by mechanical schemes will never lead to more than a chance difference between the groups on relevant variables, and since the difference for any one variable is purely chance, one cannot expect the difference to have more than a chance effect on the result for the experimental variable. If, for example, a chance, *i.e.* nonsignificant statistically, difference exists in the initial mean reading scores of two groups, this difference, in and of itself, will not lead to a significant difference in the relative extent to which they will profit from two diverse methods of teaching improvement in reading. The sampling error formula is adequate for evaluating such chance phenomena.

It should be noted here that an original group which is split into halves either at random or by pairing must be regarded as representative of some defined universe, and that such conclusions as are drawn from the experiment cannot be generalized unless it can be shown that the defined universe is representative of the generality of mankind with respect to the variable being studied. In other words, those who persist in using the college sophomore as a laboratory representative of mankind have not avoided, by showing that selective factors did not render their experimental and control groups noncomparable, the necessity of bridging the gap between the sophomore's behavior and that of the typical human.

It is interesting to consider the use of experimental and control groups in the light of a variation in the standard error of the difference between means as given in a paper by Wilks (53). Let x be the variable under study and y a possible variable for control; then, if the individuals of each group have been so selected as to yield identical distributions on the matching or control variable, y , the standard error of the difference between the means on the x variable will be given by

$$(H) \quad \sigma_D^2 = (\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2)(1 - r_{xy}^2)$$

where r_{xy} is the correlation between the experimental and control variables. If the matching has been made on the basis of several control variables, the given correlation becomes the multiple corre-

lation between the experimental variable and the matching variables. There are two important aspects of Wilks's mathematical contribution with which psychologists should be acquainted. The first of these is that the standard error of the difference can also be written in the form

$$(I) \quad \sigma_D^2 = \sigma_{\bar{x}_1}^2 (1 - r_{xy}^2) + \sigma_{\bar{x}_2}^2 (1 - r_{xy}^2)$$

from which we deduce the following important fact: Where two groups have been separately matched as to distribution on the same control variable, the standard error of the difference can be obtained without the restriction of the ordinary procedure, which requires that there be an equal number of cases in the two groups. This holds true, also, when several control variables have been used. The reader will note that either term in the above formula, (I), is, as might be expected, identical to formula (D) given earlier in this paper for the sampling error where the stratified method is used. Formula (I) is particularly useful when the cost per case in the experimental group is much greater than in the control group. Precision can be secured by taking a larger control group, a procedure which can, of course, be followed if the groups are not equated except by randomization.

The second significant fact about the Wilks formulation seems not to have been implied by him or noticed by others. When the number of cases in the experimental and control groups are equal and the distributions on the control variables for the two groups have been matched *or* the groups have been equated by pairing on the basis of these same control variables, it can be shown algebraically that the correlation, in formula (G), between pairs on the experimental variable is equal to the square of the correlation in the Wilks formula, (I). Thus, formula (G) may be written in the form

$$(J) \quad \sigma_D^2 = \sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2 - 2r_{xy}^2 \sigma_{\bar{x}_1} \sigma_{\bar{x}_2}$$

where r_{xy} is not the correlation between pairs on the experimental variable, but the multiple r between the experimental and control variables, or the zero-order correlation between the experimental and control variables when the groups have been equated on only one variable.

Formula (J) makes explicit what we have already said—namely, that the control variable or variables must be related to the experimental variable in order that the equating of groups by pairing or matching result in a statistical, hence experimental, advantage.

Furthermore, the efficacy of using additional controls is somewhat limited by the well-known fact that the increase in the multiple correlation coefficient resulting from adding more variables is usually slow. That this phenomenon of diminishing returns, associated with the problem of multiple correlation, should be operative here has probably not been suspected by experimentalists.

This multiple r must be .866 to diminish the sampling error by one-half, and .707 to lead to a reduction in error equivalent to that obtained by doubling the size of the samples. It is not our purpose to discourage the practice of equating experimental and control groups, but we do feel that investigators should realize that such procedures do not always lead to any marked advantage over the random method. In so far as greater precision can be obtained and selective factors avoided, the equating of groups does justify the personal satisfaction of knowing that the groups are comparable with regard to certain characteristics. It must be remembered, however, that, despite the matching of pairs on some variables, there are likely to be other variables of equal importance upon which the groups will be no more comparable than expected on the basis of randomization.

The writer finds himself in marked disagreement with some of the propositions set down by Corey (4) in a recent paper, entitled "The dependence of chance factors in equating groups." He attempted an empirical check on the operation of chance by means of shuffling 1000 cards with scores thereupon and drawing (but not replacing) 25, 50, 75, and 100 cards. The computed means showed the expected scatter. For $N = 25$, the difference between the extreme means was 11, with a standard error of 4.9. Analytical consideration would have foretold this finding. Corey concludes therefrom "that the practice of using various sections of the same courses in psychology for experimental and control groups is unsound." He argues that selective factors enter to make section groups noncomparable, but he fails to state just *how* such selection takes place. He also claims that the difference between section groups with $N = 25$ will be larger if the sections are recruited from smaller total groups. The reverse of this statement happens to be true, as can be seen by examining the sampling error formulas for samples drawn from a finite universe. As to the inadequacy of such small groups as 25, Corey could very easily, on the basis of his method, prove (?) that subsamples of 1000 are not comparable, since

just as significant (statistically) differences would arise in this case as when comparing subsamples of 25 each.

Rulon (37) has recently developed a new method for equating groups without the necessity of pairing or matching, but no sampling error formulas were presented to accompany the technique. Obviously, the ordinary formulas are inapplicable, since the principle of simple sampling is disturbed by the purposive selection or elimination of cases.

Our discussion is applicable only to those studies wherein groups are being compared as to central tendency (and variability), but the principle of control by matching can be used in studies involving other statistical measures, even though the necessary mathematical formulations of the resulting sampling errors have not yet been derived. The research on the nature-nurture problem by Leahy (24) not only illustrates the use of experimental and control groups in a correlational study, but also provides one of the best available examples of a research project which was well planned from both the sampling and experimental viewpoints. We state here in brief the essentials of her plan. The experimental or foster-child group was limited to (1) children placed in their adoptive home at the age of six months or younger, (2) foster children and adoptive parents of white race, non-Jewish, north European extraction, (3) ages of 5-14 at the time of testing, (4) residence in communities of 1000 or more, and (5) children legally adopted by married persons. For a control (own child) group, each adopted child was matched with an own child for (1) sex, (2) age (within plus or minus six months), (3) father's occupation, (4) father's education, (5) mother's education, (6) race—white, non-Jewish, north European extraction, and (7) residence in communities of 1000 or more. Thus, the two groups were chosen so as to rule out possible selective factors and to render the groups comparable on factors which might affect the parent-child resemblance as measured by the correlation coefficient. In this case there is no existing statistical technique for making allowance for the fact that the two groups were not independent samples; the direction of such a correction would, however, operate so as to reinforce her generalizations.

Let us turn next to a couple of monographs in which matching took place and no statistical allowance was made for this fact, even though involving nothing more complicated than formula (G). In a recent study (5) of emotional differences of delinquent and non-delinquent girls of normal intelligence, an elaborate scheme of pairing

was utilized to make the groups comparable with regard to age, IQ, cultural (home) environment, and occupational status of father, but in making the final comparison of the two groups, so matched, no account is taken of the correlational term in the sampling error formula. If, as argued, the factors used in pairing were really important, an appreciable pair correlation on emotion should have resulted. If no r existed, we have proof that the matching variables were of no consequence in the study—that is, that experimental control of irrelevant variables was useless.

In a study of attitude and unemployment (13), a group of employed engineers was chosen so as to be comparable with a group of unemployed engineers as regards age, salary (when unemployed were employed), nativity, education, religion, state licensing, and marital status. Ignoring the fact that the error formula used was inadequate, let us ask whether this apparently well-planned investigation really avoided selective factors. The thesis is that unemployment affects the attitudes of men. This might very well be so, but in concluding that the found differences support this contention, the likely selection as regards attitude and personality characteristics which may lead to one engineer retaining, another losing, his job is ignored. This investigator is also guilty of attempting to check the adequacy of his samples by splitting the groups into halves and comparing the means for halves.

As is well known, one of the most efficient experimental designs is the use of the individuals of a group as their own control. The performance of a group of individuals is determined for two different experimental conditions, and the resulting change, increase or decrease, in the behavior is interpreted as being due to the differences in conditions, provided such factors as practice effects, fatigue, memory, etc. have been taken into account. From the sampling viewpoint, such a setup does not involve the question of the comparability of two groups, but the individuals used must be regarded as a sample of some definitive universe, so that the end result must be evaluated in terms of sampling in order to have some estimate as to the likely fluctuation which would occur if the experiment were repeated on another sample of the same size. In fact, a difference is again being tested for its nonchance significance, and the standard error for doing this can be obtained in either of two ways: by computing the mean of the distribution of differences (changes, increases or decreases) and its standard error by dividing the standard deviation of this distribution by the square root of N ; or by determining

the mean performance for each condition, the difference between the means (equals the mean of the difference), and the standard error of the difference by formula (G). The latter procedure involves computing the correlation between the two sets of performances; this correlation may or may not be of interest *per se*.

This procedure of allowing a group to be its own control is in common use in laboratory work and in certain problems in educational psychology. That the procedure can be used to advantage in social psychology, particularly in the study of attitudinal behavior, is well illustrated by the research of Thurstone (49) on the changes in attitude toward the Chinese brought about by seeing the movie "Sons of the Gods."

Another method of securing comparable groups is to select control individuals who are consanguineous to the individuals in the experimental group. This includes the split-litter technique, the use of siblings as controls, and the method of co-twin control. In so far as the variation in the experimental variable is influenced by genetical and environmental factors, the use of identical twins represents the best possible method of securing comparable groups for experimental purposes. The possible advantage of using twins in one field of investigation has been pointed out by "Student" in his 1931 paper (46) on the Lanarkshire milk experiment in England. This investigation involved the daily feeding of 5000 children three-fourths of a pint of raw milk and another group of 5000 an equal amount of pasteurized milk over a period of four months. These 10,000, plus a control group of 10,000, were measured at the beginning and end of the four-month period for height and weight. Despite large numbers, the groups were not comparable as regards initial height and weight, the operating selective factor being the benevolent attitude of school teachers who apparently thought the research project would not be harmed if preference was given frail, undernourished children in choosing individuals for the feeder groups. Either a carefully supervised random, or a definite pairing, procedure would have, of course, avoided this selective bias, but what is more important and more relevant to our present topic is "Student's" claim, so far not refuted, that the use of 50 pairs of identical twins would have yielded as precise information at only 2% of the cost of the original experiment, or at a saving of approximately \$35,000.

THE SINGLE CASE

One issue involving experimental methodology which is somewhat perplexing to the statistically minded is that pertaining to the

use of a single case. The statistician who fails to see that important generalizations from research on a single case can ever be acceptable is on a par with the experimentalist who fails to appreciate the fact that some problems can never be solved without resort to numbers. The single-case method and the statistical method are, of course, somewhat opposed, but each has its merits and each its shortcomings. Many examples could be enumerated in which a single case provides sufficient data for checking hypotheses and drawing generalizations.

The writer is in no position to debate the pros and cons of the single-case method. Whether more than one individual should be studied will depend upon the nature of a particular research. The essential considerations which need to be kept in mind would seem to boil down to two: Does the behavior characteristic being studied vary greatly from individual to individual, or is this variation so slight in terms of the experimentally produced variations that the factor of individual differences can be ignored? If one does not have some knowledge of interindividual variation, it may be necessary to use several cases to demonstrate its presence or absence. When the single case is made available by accidental factors, it is not always possible to use more than a single case, nor is it necessary to do so, providing the research is dealing with relatively nonvariable behavior and providing the results from the single case fit in with a large number of other established facts or a number of carefully conceived hypotheses.

From the statistical viewpoint, a single case can always be taken as representative of the generality of mankind when the investigator is dealing with behavior or responses which do not show individual differences. There is no sampling problem here, but it becomes the responsibility of the investigator to show either that variation does not exist or that such changes as are produced or observed are greater than any possible individual differences. Some psychologists have claimed that a single *ideal* case is sufficient for scientific purposes, but the realization of adequate experimental controls is so far from ideal that such a concept seems to be a flight from reality. If there is such a thing as general psychology—a science which learns what is true for individuals in general—then such a science could be built upon a single case with no thought to qualifications for individual differences. The use of several cases and any of the various types of statistical averages may or may not lead to a general psychology—such averages may not only cause us to lose sight of the individual variation, but may also mask a fact of pertinence to

general psychology; the plateau as a general characteristic of so many types of learning can easily be lost in the process of averaging.

In the field of personality research one finds that the intensive study of one individual is being advocated. It is apparently thought that complete knowledge of one person is better than incomplete information on large numbers. This may be true so far as the one individual is concerned, but one can very well raise a question as to the possibility of generalizations concerning the behavior of others. One may also wonder what reference points are used in evaluating the many observations on the personality of a single individual. It does not seem unreasonable to suppose that some reference point, other than the subjective orbit of the investigator, is a requisite for an objective science of behavior. To claim that the subject provides his own frame of reference *via* patterns of responses may sound like a way of escaping the dilemma, but it is difficult to see what can be said regarding the personality of men in general by following the pattern for one individual. As the number of behavior characteristics observed is increased, the more complex becomes the pattern and the greater the possibility of the pattern itself showing individuality. Surely, psychologists have learned that very little light is thrown on, say, criminal behavior by a minute clinical study of one case, yet we are expected by some to believe that the mysteries of human personality will somehow be unraveled by an intensive study of just one case. Perhaps knowing all about one case may be important, even though of highly limited significance for the next and the next case.

SUMMARY AND CONCLUSIONS

When the search for material for this paper was first begun, it was thought that a useful critique of sampling as used in *social psychology* could be prepared. It was soon discovered that specific sampling techniques were so few that such a critique would not be justified, but further search revealed that certain general principles of sampling, somewhat unfamiliar to psychologists, could be gleaned from the rather widely scattered literature of statistical methodology. It was apparent that such general principles and specific sampling techniques as have been set forth were applicable not only in *social psychology*, but also in other fields of psychology. Hence, the present paper has been prepared with a wider horizon of usefulness in mind than originally planned.

We have attempted to clarify certain aspects of the sampling problem and have given an account of the present status of sampling

methods. The reader who searches the foregoing pages cannot be any more disappointed than the writer concerning the paucity of techniques for drawing and checking a representative sample. Since sampling is so basic to much of the research in psychology, investigators should seize every opportunity to investigate this adjunct of research. In particular, samples drawn by different methods should be checked, the one against the other. This is especially important in field studies.

Certain aspects of the problem have been treated too briefly in the foregoing pages. The question regarding the requisite size of a sample or samples needs further elaboration. The degree to which distributions may be skewed without disrupting the sound use of standard errors in judging the significance of statistical results is in need of clarification. The concept of homogeneity should be more fully discussed. Some of the questions raised concerning the choice of individuals for experimental and control groups should receive further consideration. The applicability in psychology of certain of Professor R. A. Fisher's designs should be examined. Eventually, the analysis of variance will come into use in psychological research; an expository paper thereon would not be without value.

In closing this paper, the writer is inclined to agree with the skepticism expressed by Bowley (3). This skepticism is based upon the existing ignorance concerning the adequacy of the available techniques and is bolstered by the not infrequent flukes of sampling which simply cannot be ascribed to chance. Perhaps the confidence to be placed in the results of a study should vary directly with the amount of information concerning the sampling and experimental techniques rather than inversely with the square root of the number of cases.

BIBLIOGRAPHY

1. BOWLEY, A. L. *Elements of statistics*. New York: Scribner, 1926.
2. BOWLEY, A. L. Measurement of the precision attained in sampling. *Bull. int. statist. Inst.*, 1926, **22**, Pt. 1, Appendix, 6-61.
3. BOWLEY, A. L. The application of sampling to economic and sociological problems. *J. Amer. statist. Ass.*, 1936, **31**, 474-480.
4. COREY, S. M. The dependence upon chance factors in equating groups. *Amer. J. Psychol.*, 1933, **45**, 749-752.
5. COURTHIAL, A. Emotional differences of delinquent and non-delinquent girls of normal intelligence. *Arch. Psychol., N. Y.*, 1931, **20**, No. 133.
6. CROSSLEY, A. M. Straw polls in 1936. *Publ. Opin. Quart.*, 1937, **1**, 24-35.
7. EZEKIEL, M. "Student's" method for measuring the significance of a difference between matched groups. *J. educ. Psychol.*, 1932, **23**, 446-450.

8. EZEKIEL, M. Reply to Dr. Lindquist's "further note" on matched groups. *J. educ. Psychol.*, 1933, **24**, 306-309.
9. FISHER, R. A. *The design of experiments*. London: Oliver & Boyd, 1937.
10. FRANZBLAU, R. N. Race difference in mental and physical traits. *Arch. Psychol., N. Y.*, 1935, **26**, No. 177.
11. GARRETT, H. E. *Statistics in psychology and education*. New York: Longmans, Green, 1937.
12. GARRETT, H. E., BRYAN, A. I., & PERL, R. E. The age factor in mental organization. *Arch. Psychol., N. Y.*, 1935, **26**, No. 176.
13. HALL, O. M. Attitude and unemployment. *Arch. Psychol., N. Y.*, 1934, **25**, No. 165.
14. HANNA, H. S. Adequacy of the sample in budgetary studies. *J. Amer. statist. Ass.*, 1934, **29** (Suppl.), 131-134.
15. HILTON, J. Enquiry by sample: an experiment and its results. *J. roy. statist. Soc.*, 1924, **87**, 544-570.
16. JENSEN, A. Report on the representative method in statistics. *Bull. int. statist. Inst.*, 1926, **22**, Pt. 1, 359-378.
17. JENSEN, A. The representative method in practice. *Bull. int. statist. Inst.*, 1926, **22**, Pt. 1, 381-439.
18. JENSEN, A. Purposive selection. *J. roy. statist. Soc.*, 1928, **91**, 541-547.
19. JOHNSON, D. A., & EURICH, A. C. An empirical test of sampling. *J. exp. Educ.*, 1935, **3**, 174-179.
20. JONES, H. E., & CONRAD, H. S. The growth and decline of intelligence: a study of a homogeneous group between the ages of ten and sixty. *Genet. Psychol. Monogr.*, 1933, **13**, 223-298.
21. KATZ, D., & CANTRIL, H. Public opinion polls. *Sociometry*, 1937, **1**, 155-179.
22. KISER, C. V. Pitfalls in sampling for population study. *J. Amer. statist. Ass.*, 1934, **29**, 250-256.
23. KLINEBERG, O. A study of psychological differences between "racial" and national groups in Europe. *Arch. Psychol., N. Y.*, 1931, **20**, No. 132.
24. LEAHY, A. M. Nature-nurture and intelligence. *Genet. Psychol. Monogr.*, 1935, **17**, 235-308.
25. LINDQUIST, E. F. The significance of a difference between "matched" groups. *J. educ. Psychol.*, 1931, **22**, 197-204.
26. LINDQUIST, E. F. A further note on the significance of a difference between the means of matched groups. *J. educ. Psychol.*, 1933, **24**, 66-69.
27. LURIE, W. A. A study of Spranger's value-types by the method of factor analysis. *J. soc. Psychol.*, 1937, **8**, 17-37.
28. MANGUS, A. R. Sampling in the field of rural relief. *J. Amer. statist. Ass.*, 1934, **29**, 410-415.
- ✓ 29. McCORMICK, T. C. Sampling theory in sociological research. *Social Forces*, 1937, **16**, 67-74.
- ✓ 30. MELTON, A. W. The methodology of experimental studies of human learning and retention: I. *Psychol. Bull.*, 1936, **33**, 305-394.
31. NEYMAN, J. On two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *J. roy. statist. Soc.*, 1934, **97**, 558-606.
32. NEYMAN, J. Contribution to the theory of sampling human populations. *J. Amer. statist. Ass.*, 1938, **33**, 101-116.

33. PEATMAN, J. G. Hazards and fallacies of statistical method in psychological measurement. *Psychol. Rec.*, 1937, 1, 365-390.
34. ROBERTS, C. S. Ordinal position and its relationship to some aspects of personality. *J. genet. Psychol.*, 1938, 53, 173-213.
35. ROBINSON, E. A. Trends of the voter's mind. *J. soc. Psychol.*, 1933, 4, 265-284.
36. ROSS, F. A. On generalization from limited social data. *Social Forces*, 1931, 10, 32-37.
37. RULON, P. J., & CROON, C. W. A procedure for balancing parallel groups. *J. educ. Psychol.*, 1933, 24, 585-590.
38. SCHILLER, B. A quantitative analysis of marriage selection in a small group. *J. soc. Psychol.*, 1932, 3, 297-319.
39. SCHOENBERG, E. H., & PARTEN, M. Methods and problems of sampling presented by the Urban Study of Consumer Purchases. *J. Amer. statist. Ass.*, 1937, 32, 311-322.
40. SMITH, J. G. Elementary statistics. New York: Holt, 1934.
41. SORENSEN, H. Mental ability over a wide range of adult ages. *J. appl. Psychol.*, 1933, 17, 729-741.
42. SORENSEN, H. Statistics for students of psychology and education. New York: McGraw-Hill, 1936.
43. STECKEL, M. L. Intelligence and birth order in family. *J. soc. Psychol.*, 1930, 1, 329-344.
- ✓44. STEPHAN, F. F. Practical problems of sampling procedure. *Amer. sociol. Rev.*, 1936, 1, 569-580.
45. STOUFFER, S. A. Statistical induction in rural social research. *Social Forces*, 1935, 13, 505-515.
46. "STUDENT." The Lanarkshire milk experiment. *Biometrika*, 1931, 23, 398-406.
47. SUKHATME, P. V. Contribution to the theory of the representative method. *Roy. statist. Soc. Suppl.*, 1935, 2, 253-268.
48. TERMAN, L. M., & MERRILL, M. A. Measuring intelligence. New York: Houghton Mifflin, 1937.
49. THURSTONE, L. L. The measurement of change in social attitudes. *J. soc. Psychol.*, 1931, 2, 230-235.
50. TIPPETT, L. H. C. The methods of statistics. London: Williams & Norgate, 1937.
51. WALKER, H. M. The sampling problem in educational research. *Teach. Coll. Rec.*, 1939, 30, 760-774.
52. WILKS, S. S. The standard error of the means of "matched" samples. *J. educ. Psychol.*, 1931, 22, 205-208.
53. WILKS, S. S. On the distribution of statistics in samples from a normal population of two variables with matched sampling of one variable. *Metron*, 1932, 9, 87-126.
- ✓54. WOOFTER, T. J. Common errors in sampling. *Social Forces*, 1933, 11, 521-525.
55. YATES, F. Some examples of biased sampling. *Ann. Eugen., Camb.*, 1935, 6, 202-213.
56. YULE, G. U. An introduction to the theory of statistics. London: Griffin, 1929.

MENTAL MEASUREMENTS IN PRIMITIVE COMMUNITIES

BY CECIL WILLIAM MANN

University of Denver

INTRODUCTION

For a number of reasons there has been no satisfactory solution to the problem of race differences. The intrinsic difficulty of the problem, the inadequacy of the methods and the instruments employed, and the emotional bias attached to the concept of white race superiority have all been obstacles in the solution of the problem. There have not been wanting, however, any number of opinions relative to the nature and amount of race differences; opinions based upon rationalizations, preconceptions, and prejudice, or upon scanty and incomplete measurement.

In some respects, the quest for a solution has resembled the classic game of "passing the buck." During the early part of the Nineteenth Century, clergymen and others, impressed by the obvious differences in the physical appearances and in the customs of races, and feeling the need for a justification of slavery as a social institution, rationalized that these differences were innate and produced indubitable evidence in favor of the superiority of the white race. Priest's assertion (81) that the inferiority of the Negro was the result of Noah's curse of perpetual slavery upon Ham, who saw the nakedness and intoxication of his father, was but one of a number of ingenious Biblical arguments invented by theologians and accepted by laymen.

Because they tired of the game, or perhaps because they had exhausted their ingenuity in rationalization, the theologians passed the buck to social philosophers and anthropologists, confident that the latter could, or would, find no evidence which would shake their claims for the inalienable superiority of the white race. Handicapped by this theological bias and lacking refined measuring devices, it is not to be wondered that some social philosophers were soon producing as much armchair evidence as did the theologians for innate race differences. Even what might have been regarded as objective evidence became the victim of the rationalizations of the protagonists of white superiority. When Bache (5) found Indians fastest and whites slowest in reaction times, it was argued by some that even here the

whites were superior in that they had developed a greater capacity for inhibition.

On the other hand, but few anthropologists were influenced by race prejudice. Anthropological evidence was not lacking to suggest that non-whites were equal in some and superior in other traits to whites. Moreover, most anthropological writings adhered to the 'psychic unity of man.' More concerned with race differences than with superiority, Kroeber (42) suggests that it is a "difficult task to establish any race as either superior or inferior, but relatively easy to prove that we entertain a strong prejudice in favor of our own racial superiority" (p. 85). Even more forcefully, Goldenweiser writes:

"As one becomes immersed in the study of racial psychology one comes to realise that the significant factor involved is not by any means the psychical differences of the races, but rather the psychical unity of man. . . What counts and demands attention is not the problematical difference in racial ability, but the disability of the genus *Homo*, however *sapiens*, to think intelligently and without prejudice in this field, so heavily charged with emotion, vanity, special pleading and still lowlier affects" (31, p. 36).

It is significant, too, that interest turned more to the field of social anthropology, leaving many of the problems of mental differences to the psychologists. Once again the buck had been passed.

Showing little reluctance—on the contrary, with a good deal of zest and enthusiasm—the psychologists took up the problem, and either because they were less astute or more tenacious have stayed with it ever since. The development of mental tests in the first decade of this century and their subsequent rapid spread made it now appear that here was an easy and unequivocal method of solving the problem once and for all time. The literature of the past 30 years is full of the results of comparative tests of racial differences.

It is likely that the psychologists would have solved the problem—at least to their own satisfaction—but for the inconvenient, yet necessary, warning of the statisticians, with their insistence upon adequate selection of groups, descriptions of sampling errors, and other statistical checks. Instead of a speedy solution of the problem, we have been forced to return to a more careful consideration of methods and techniques. It would appear, indeed, that a solution of the problem by the use of the methods of the past 30 years is difficult, if not impossible. Slowly, but surely, the psychologists are, in turn, passing the buck. This time, it will probably fall into the laps of the statisticians interested in the problems of psychology.

THE PROBLEM OF RACE DIFFERENCES

Races are different. From the casual observations of travelers, the studies of the anthropologists, and the more or less carefully controlled investigations of the psychologists, sufficient evidence has accumulated to warrant this statement. The problems facing the investigator of race differences are those of determining qualitative and quantitative measures of differences and of estimating the influences which have produced the differences. The determination of race differences is an extension of the nature-nurture problem. It is concerned with investigations of the operation of the hereditary and environmental factors in the variation of the individual, examination of the conditions—if any—under which it is possible to modify these factors, and the prediction of the results of such modification.

In order to clarify the position, it will be necessary to consider the two main groups of investigations which have generally been included in the scope of race measurements. In the first place, there is a large amount of material in which direct and indirect measurements have been made of physical and psychophysical traits. In this group are placed the measurements of height, weight, cranial capacity, sensory acuity, and the like. The second group consists of the investigations of mental traits by the use of tests which are by nature sampling rather than measuring devices. In this group we should place the many investigations which have attempted to measure general ability, specific abilities, and temperament.

The arguments for and against the classification of individuals by the use of sampling devices are as valid for primitive groups as they are for members of more complex communities. The technique is open to serious objection, however, when the samples of one culture are compared without qualification by devices which are but measures of the samples of another culture. Particularly is this true when attempts are made to compare *innate* differences between individuals or groups of different cultures by devices which are generally accepted as being valid only in the culture in which they were constructed and standardized. Under such conditions of investigation we shall certainly find differences, but, by them, we can establish nothing beyond the already-known fact that cultures are different. We shall have failed to touch the problem of innate differences.

The use of sampling devices, such as tests of general and specific abilities, may be the basis of at least two measures. In the first place, there is an obvious value in the use within a culture of a sampling

device which will differentiate the mental abilities and traits of the members of that culture. In the second place, under carefully controlled conditions—conditions which have rarely, if ever, been realized in past investigations—there may be the possibility of interpreting differences of comparable samples in terms of race differences. If any progress is to be made in the direction of the discovery of innate differences in different culture groups, it will be necessary to control the culture variables, to establish valid measuring or sampling devices, and to interpret the results in terms of the control and devices that have been used.

The conditions which ought to be met by those investigating race differences have been emphasized by Arlitt (4), Blackwood (9), Daniel (16), Freeman (25), Goodenough (33), Mead (63), Oliver (67, 72), Peterson (74), and others with, one would have thought, sufficient clarity and emphasis to make further restatement unnecessary. However, in view of the fact that an evaluation of the studies which form the basis for this review is to be made in terms of the measuring conditions which ought to be met, it will be necessary to mention briefly some of the special difficulties which lie in the way of valid race comparisons.

Control of Cultural Variables

Language. This is one of the most obvious as well as one of the most important obstacles in the way of race comparisons. Comparisons of races on the basis of tests in which any of the groups compared suffer the slightest handicap because of language differences must be excluded from the evidence for or against innate differences.

Physical Environment. Familiarity with the elements which form the test items will be one measure of the worth of a test. The form of the present intelligence test attempts to include samples of performance which are representative of the culture in which it is to be used. For obvious reasons, the samples used must be limited in number. The best tests are those in which the samples used are most representative of the culture and most discriminative of the trait that is being measured. If any degree of reliance is to be placed upon "international" tests which purport to measure race differences, it must have been established that the items which comprise the test are as truly representative of, and as truly discriminative in, any one culture as in any other culture. In other words, we should demand of an "international" test that it be as valid and reliable in every culture in which it is used as is the best mental test in use in any one

culture. The obstacles in the way of an "international" test are obvious. One has but to consider some of the different customs in different cultures with respect to such universals as birth, death, marriage, the use of one's name and the names of relatives, movements of the solar system, climate and the seasons, to realize the responsibility attached to the construction of an "international" test.

Social Environment. Valid race comparisons will involve the due consideration of the essential differences in social environment. Few investigations, for instance, have made adequate allowance for the differences in attitude in individualistic and communalistic societies. In a society based upon the latter system, where chiefdom is part of the social fabric, it may be impossible for an individual to move away from his born status. Wherefore, social expectancy will tend to operate against the kind of individual competition which forms the basis of all test situations. In communalistic systems it has been reported many times that the natives were greatly surprised that they were not allowed to assist each other in tests. Faced with a new difficulty and denied the usual assistance of their friends, they soon lost interest in the test.

Many comparisons of groups of mixed-bloods have not always taken into account the social status of the group. Often these people are at a serious disadvantage because of their social position. Frequently the half-caste is an outcast. Unless we can be certain that the investigator is aware of the social status of the mixed-blood and has controlled his investigation accordingly, we must be skeptical of the results.

Other factors such as the halo effect of the foreign investigator, the difficulty of obtaining complete 'rapport,' the cultural attitude toward time, motivational factors—these and many other elements must, of necessity, be controlled if valid comparisons are to be made.

Selection of Groups

Comparisons of individuals or of groups are valid only when such individuals or groups are comparable. Unless we can be certain that the "population" being measured is representative of the general population we have no sanction for applying the results gained from the "population" to the general population. The results gained by the measurement of a "population" which is representative of the general population in some respects may not be applied to the latter unless we can be certain that the variables under consideration are

not influenced by elements in which the "population" differs from the general population.

Age. In many primitive communities, birth registration is but partly controlled, and birthdays pass unnoticed. The selection of groups on the basis of age in such communities must be viewed with the greatest caution, and the comparisons made accepted with reserve.

Socioeconomic Status. The application of any of the well-known scales for measuring the social and economic conditions of primitive peoples would be certain to reveal great deficiencies, but would give scarcely any indication of the true status of the individuals. Yet, if comparisons are to be valid, the socioeconomic status of the groups as a factor in the production of variability must be controlled in comparative investigations.

Length of Schooling. The difficulty of ascertaining the length of schooling is comparable to that of discovering age. Even if the records are accurate—and this is by no means a certainty—it would be folly to assume that four years of schooling in Alaska or in Fiji are the equivalent of the same length of schooling in America.

Culture-Contact. The influence of the quantity and quality of the culture-contact in primitive groups is as easy to overlook as it is difficult to estimate. The task of equating the differential influences of culture-contact among the partially controlled New Guinea natives, the Australian aborigines, between Fijians and Hawaiians, is one that should make even the most daring investigator somewhat nervous of his comparisons. Even more hazardous will be the task of estimating the influences of culture-contact when a group migrant from one culture is placed under the influences of more than one culture, all of which are different.

Selective Immigration. Comparisons of race groups alien to a culture will be valid only for the groups compared, and extension of the conclusions to the parent group cannot be justified. It is obvious that the circumstances which have determined the emigration may well be a factor in determining the character of the group which emigrates.

It would not be difficult to enlarge upon all the difficulties which have been mentioned and to include others. One's conviction of the validity of race comparisons will rest upon one's belief that in the investigation there have been taken into account all the possible variables—language, environment, culture-contact, and the like—which may be productive of test score variables.

During the present decade there has been considerable activity in

the field of race measurement. This review will be limited to the investigations of mental differences made since 1930 and to the areas bordering the Pacific Ocean and to Africa. The bibliography has been enlarged to include investigations which have been made in the above areas with respect to special senses (11, 38, 89), reaction time (34), perception and learning (39, 51, 93), eidetic imagery (73), and artistic (2, 64) and musical (18, 69, 83) abilities.

ALASKA

One of the most extensive investigations of primitive culture in terms of objective tests was that of Alaskan natives made by Anderson and Eells in 1930-1931 (3, 20, 21, 22). The measurements were but one part of the total program, which included investigations of the socioeconomic and educational status of the natives of Alaska. A total of 1084 children were tested with from 1 to 23 tests, and a total of 13,724 measures were obtained. The program included tests of general, physical, musical, and mechanical abilities, and general tests of scholastic achievement, handwriting, English composition, and musical accomplishment. It is believed by the investigators that the results obtained are some indication of the variability within the races tested and may, with qualifications, form the bases of comparison between Alaskan natives and whites.

An effort was made to secure a representative sample of the school population in terms of the total school population, purity of blood, type and size of schools, geographical position, grade placement, age, and ability. That their efforts in this direction were but partly successful was due to the difficulties of travel in an arduous climate rather than to their insensibility to the serious limitations which faulty sampling would place upon the interpretation of the results.

Measurement of General Ability

In an effort to secure an adequate classification of the school population two tests were used as measures of general ability, the Stanford-Binet Revision (1916) and the Goodenough test of drawing a man.

The Stanford-Binet Scale. An investigator with training and experience in the use of the Stanford-Binet scale accompanied the party and administered the test, perforce in English. A little experience with the test in Alaska soon convinced the investigators that modifications of the tests were necessary. The modifications con-

sisted of the substitutions or omissions of items due to verbal difficulties or cultural differences, *e.g.* substitution of "captain," "fish," and "boat" for "engineer," "car," and "train," and substitution of "muskrat" for "snake," "reindeer" for "cow," "duck" or "ptarmigan" for "sparrow." In all, 11 tests were omitted; in three cases alternate tests were used; in four year-groups less than six tests were used, while approximately a dozen words or phrases were substituted (3, p. 306). "It was felt that these (modifications) tended to make it a fairer measuring instrument for the desired purpose under the conditions which obtained in Alaska" (3, pp. 306-307).

Were the results to be used only for the purpose of making a classification of the children of Alaska, one would agree that the alterations were justified and might go still further and suggest others. When, however, the results are to be used, as they were, as the basis of a comparison of Alaskan and American children, one has a right to know the effects of the modifications upon the standardization of the test, and, failing this knowledge, to view with caution any comparisons which are made.

The scores obtained were transformed into intelligence quotients. The results are shown in Table I.

TABLE I
MEAN IQ'S FOR ALASKAN NATIVE RACES AS MEASURED ON THE STANFORD-BINET SCALE

Race	N	Mean IQ	σ	IQ's Exceeding 100	
				N	%
Eskimo	389	73.67	12.78	16	4.1
Aleut	94	80.27	13.94	5	5.3
Indian	83	78.98	14.92	6	7.2

It will be seen that there is a significant difference between the mean IQ's of Aleuts and Indians and that the mean IQ of the Eskimos is significantly lower than those of the other races.

A distribution of IQ's according to sex was made, and slight differences were observed. Only in the case of the Eskimos were the differences significant, and it is suspected that this difference was due to some undetermined selective factor.

It is necessary to point out that the test was administered in English. There are at least 18 dialects used among the Alaskans, and no common language. Obviously, there will be a major language handicap.

"... 76.8 per cent (are) unable to read and write the English language and (are) dependent upon their own tongue for social com-

munication. . . Nor is bilingualism pronounced. Children learn and use English in the schoolroom, but as soon as they leave the schoolroom they are back in their native language environment. Teachers have exerted herculean efforts to break the force of the tribal tongues, but to no avail. The Eskimo language is still the dominant means of expression, and English is only something needed for communication with whites" (3, p. 189).

In his study of the American-born Japanese, Darsie (17) concluded that the mean IQ was spuriously low, not because of inferior innate ability, but rather because of the language difficulty. His findings suggest that the mean IQ for Japanese children of 13 years of age may be as much as eight points too low. Bell's findings are essentially of the same order (6). Walters estimates (94) that the language handicap, regardless of innate ability, may account for a retardation of from six to eight months of mental age for children of 13 years of age coming from foreign-speaking homes in New York.

TABLE II
MEAN IQ'S FOR ALASKAN NATIVES USING GOODENOUGH SCALE

	Eskimo	Aleut	Indian
N	364	105	58
Mean	89.56	93.29	91.55
σ	15.57	15.59	13.74

The authors of the Alaska study suggest that the deficiency in IQ due to a language handicap may be from two to five points. It should be noted, however, that while the foreign children measured by Darsie, Bell, and Walters spoke foreign languages, they lived in an English-speaking community and thereby had more opportunities for incidental learning of English than have the Alaskans, whose only English-speaking contact is in school. The deficiency due to language handicap would, it is likely, be no less in Alaska than it would be in America.

The Goodenough Scale. In order to secure another measure of mental ability the Goodenough scale of measuring ability by drawing a man was used (32). The results are shown in Table II. Comparing these results with those obtained on the Stanford-Binet scale (Table I), it will be noticed that there are marked differences, all of which are significant and in favor of the Goodenough scale. These are shown in Table III.

It is suggested that "possibly 15 points of IQ may be due to the language factor involved in the more reliable Stanford-Binet" (3,

p. 313). It must be pointed out, however, that the Stanford-Binet is reliable only over the group on which it was standardized or a comparable group. As a measure of Alaskan mental ability it is probably no more reliable than would be an Alaskan test of mental ability standardized on an Alaskan group and then applied to a group of Californian children. It was this notion of comparable measuring devices which was responsible for the rejection of the Binet-Simon scale as an instrument for measuring the mental ability of English-speaking children, and the subsequent standardizations of tests to suit the cultures in which they were to be used.

Before leaving the problem of the measurement of general ability, the investigation deals with certain cognate problems which should be mentioned.

TABLE III

COMPARISON OF MEAN IQ'S AS MEASURED ON STANFORD-BINET AND GOODENOUGH SCALES FOR ALASKAN NATIVES

Race	Stanford-Binet	Goodenough	Difference
	Mean IQ	Mean IQ	
Eskimo	73.7	89.6	15.9
Aleut	80.3	93.3	13.0
Indian	79.0	91.6	12.6

Relation of IQ to Blood Purity.

"If it is true as the evidence previously presented seems to indicate that the intelligence quotients of the primitive races, Eskimos, Aleuts and Indians, are lower than those of whites, then its value should tend to increase progressively with additional admixture of white blood" (3, p. 319).

There are probably few who would accept the argument in the form in which it is stated. Evidence for variability of IQ with variability of blood-mixture is presented, but without much conviction. Moreover, the numbers tested are far too small to make a basis for any generalization. In another part of the report the authors indicate the special disabilities of the half-castes in both white and primitive communities, and, from one's own experience, one would be inclined to suggest that they have not magnified the situation.

Relation of IQ to Degree of Contact With White Culture.

To study the factor of white contact, the scores of full-blooded Eskimo children were grouped into "primitive," "semi-contact," and "full-contact," and an analysis was made. In no case is the differ-

ence of means significant. It is unfortunate that the numbers compared are too small to be regarded as reliable. Those tested could hardly have been regarded as typical of the race measured, and in the absence of data relative to the quantity and quality of the culture-contact one would be inclined to accept with caution the rather sweeping conclusion that "pure-blooded Eskimos have essentially the same mental ability regardless of the degree of their contact with white civilization" (3, p. 320).

Relation of IQ to School Experience and Age.

Data are presented for the Alaskan children showing the mean IQ's for different amounts of school and for age. For both Stanford-Binet and Goodenough scales there is an apparent decrease in mean

TABLE IV
MEAN IQ'S OF ESKIMO CHILDREN DISTRIBUTED BY AGE AT LAST BIRTHDAY

Age	Stanford-Binet		Goodenough	
	N	IQ	N	IQ
8	7	99.6	16	100.0
9	13	94.0	23	97.7
10	35	84.3	41	91.0
11	29	79.2	30	97.5
12	46	76.0	45	89.3
13	53	71.4	55	86.3
14	57	69.4	43	86.3
15	45	65.1	43	84.7
16	46	69.2	35	82.9
17	21	69.7	18	86.9
18	14	66.8	5	87.2

IQ with years of schooling. When, however, the groups numbering less than 30 are excluded, it will be seen that there is little difference in mean IQ's for the remaining groups. With respect to age, too, there is a distinct drop in mean IQ with increase of age, even when groups numbering less than 30 are excluded. It should be noticed, however, that the drop is far more marked in the Stanford-Binet than in the Goodenough results.

The results are shown in Table IV. This table, combined with the analysis of the percentages of those who passed individual tests at various age levels (3, p. 325), gives the best evidence of the unsuitability of the Stanford-Binet scale as a measure of the mental ability of the Alaskan natives.

From the complete analysis of percentages of those passing the tests we have selected a number of illustrations for Table V.

From the percentage table (3, p. 325) we learn that all children above the age of 10 years satisfy most of the tests in the VI-, VII-, and VIII-year groups with a criterion of mastery above 75%. The first serious drop in percentage is to be noticed in Year VIII, 6, Vocabulary Test, which was passed by only 32% of 10-year-old children. At the same time it will be noticed (Table V) that there is a gain for the older children in the same item, increasing to 61% at 16 years and over.

The list given is typical of the table (3, p. 325), and the outstanding facts are (a) that the lowest percentages are in the items which demand specific knowledge of a different culture, and (b) that while

TABLE V

A SELECTION OF THE PERCENTAGES OF ESKIMO CHILDREN AT EACH AGE LEVEL WHO PASSED INDIVIDUAL TEST ITEMS OF THE STANFORD-BINET SCALE

Age	10	11	12	13	14	15	16 and Over
N	28	36	30	57	53	52	111
VIII. 1. Ball and field.....	89	94	83	93	94	98	96
6. Vocabulary	32	11	14	46	42	45	61
IX. 3. Making change	29	28	20	46	49	58	67
X. 1. Vocabulary	4	3	3	18	12	6	27
6. Naming 60 words....	14	22	14	32	26	21	37
A. Healy-Fernald puzzle..	79	92	79	91	98	88	91
XII. 4. Dissected sentences...	6	4	6	12	12	13	25

there is a decrease in the percentages of those passing items at any one age, the percentages passing any item tend to increase at the higher age levels. One striking exception is to be noted in the results for the Healy-Fernald formboard, which is passed by 79% of the 10-year group, thereby reaching a satisfactory criterion of mastery, and by increasing percentages at the higher age levels.

The conclusion forced upon us is not that there is a decrease in mental ability as we pass from lower to higher age levels, but that the test becomes progressively unsuitable for Alaskans as we pass from the lower to the higher levels. Further evidence for this is found in the much smaller differences in mean IQ for the age groups as measured on the Goodenough scale. The sudden drop in the Goodenough IQ mean at 12 years might well be accounted for, as the authors suggest, by the fact that it was standardized on younger children and is probably not applicable to the older groups.

Physical and Mechanical Tests

Motor Ability. For the purpose of securing comparable measures of innate motor ability the Brace Scale of Motor Ability (12) was used. The scale consists of 20 "stunts," of which the following are samples:

- (1) Walk in a straight line, placing the heel of one foot in front of, and against the toe of the other foot. Take 10 steps in all, 5 with each foot.
- (2) Fold the arms behind the back. Kneel on both knees. Get up without losing the balance or moving the feet about.

No apparent difficulty was experienced in administering the test, but the results must have been definitely disappointing to the Alaskans.

"All three races show distinct inferiority to the given (American) norms, the amount of inferiority decreasing steadily with increasing age. In spite of their constant rigorous outdoor life and the necessity for physical activity, they do not exhibit nearly the physical agility, balance, control, flexibility and strength of white children of similar ages in the United States. This suggests a marked need for a more systematic programme of physical education, especially during the long winter months when outdoor activity is necessarily restricted" (3, pp. 330-331).

To others, it might also suggest a measure of skepticism as to the innateness of the motor abilities which the scale is supposed to measure. There is no evidence that the "stunts" are in any way applicable to the measurement of Alaskan motor ability. One would willingly undertake to select from a Fijian environment a number of "stunts" which appear "natural" to a Fijian, yet which would tax the motor ability of most white children. For instance, it is not uncommon in the coastal villages of Fiji to find children of three or four years of age who are able to swim; at five or six years of age, boys and girls can climb coconut trees like monkeys. Were we to apply to Alaskans or even American children a Fiji test of motor ability, we need not be surprised at the relatively low results which might be achieved.

From the results of the test of motor ability we can conclude nothing with respect to the motor ability of the Alaskans. It does indicate, however, that as much care needs to be exercised in the choice of items for measuring motor ability as we should like to see displayed in selecting items for the measurement of innate mental ability.

Mechanical Ability. For the measurement of mechanical ability

the MacQuarrie Test for Mechanical Ability (53) was applied to 591 children. From the results obtained, it is concluded that "the Aleuts and the Indians appear to be distinctly inferior to whites in mechanical ability, and the Eskimos still more markedly inferior" (3, p. 329).

It is to be remembered, however, that, while the MacQuarrie test is nonlinguistic, the directions involve language, and that the concepts involved are distinctly American. A better test for Alaskans would probably be the extent to which they can adapt themselves to the use of mechanical tools and skills which have been introduced by the white men.

Sensory Acuity

For testing visual acuity a Snellen chart was used where time and lighting conditions were favorable. A total of 371 children were tested. The results indicate no marked superiority of vision among Alaskan natives, and that 40% of Alaskan children had subnormal vision compared with 27% of white children.

A very rough measure of hearing was attempted by means of a whispering test. No reliability can be placed upon the results, for the conditions of testing were not always identical. No comparisons were made with whites.

Aesthetic Abilities

Musical Ability. As a measure of musical ability the Seashore Test of Musical Talent (85) was used, the test being limited to pitch, intensity, and tonal memory. A total of 551 children were tested, the results being averaged around the 30-40 percentiles as referred to American children.

The Seashore tests are nonlinguistic, but complex to administer. The requirement that all record the responses at the same time cannot be met by all. It is difficult, moreover, to make sure that the right comparisons are being made. The tests are long and tiring. The difficulties of administering the test to primitives can be appreciated only by one who has tried to do so. Results obtained from primitives need to be used with great caution.

Artistic Ability. No objective tests were used to measure artistic ability. Instead, judgments were made upon the bases of (1) samples of native crafts, (2) samples of drawings made under direction, and (3) samples of drawings made without direction. From their observations, the authors conclude that "while it is impossible to judge quantitatively the artistic ability of the average Alaskan natives by

such evidence as has been presented, undoubtedly it shows considerable artistic ability on the part of many native children of various ages expressed under dissimilar conditions" (3, p. 339). (There was administered, in addition, a program of educational achievement testing based extensively upon the Stanford Achievement Test. The consideration of the results of this program, however, is beyond the scope of this review.)

General Considerations

There are matters of general interest which apply to all the psychological tests administered in the Alaskan program.

Age. The presentation of mean IQ's with two places of decimals presumes a knowledge of individual chronological ages with an accuracy which is hardly likely to be found in Alaska, or, indeed, in some of the United States. We are told, for instance, that "of the 21 villages studied intimately by the writer only two possessed complete records of births, deaths and marriages going back as far as 1918" (3, p. 137). Few villages had data for consecutive years, and it is likely that the information concerning the ages of some of the children tested would be little better than a good guess. This should be kept in mind when comparisons are being made.

Sampling. We have no evidence that the samples tested were truly representative, and the authors have admitted this difficulty. As information concerning the individuals and groups tested, the results—with cautious interpretation—may have considerable value, but there they must rest for the present. All that we can say comparatively is that within the inadequacies of the instruments and the methods used—and these are not inconsiderable—some Alaskans measured lower than a representative group of American children. We have, from the results, no indubitable proof of the superiority or of the inferiority of the mental ability of the Alaskans.

Educational Implications. More doubtful, but with more far-reaching consequences, is the implication that, since Alaskans are inferior in a test designed for American children, their educational system should be modified in accordance with that measurement.

"As already pointed out, however, it is the practical question of the present actual ability of these races which is of chief concern in this study, for its purpose is to improve the school system organized by the Federal Government for the benefit of the Alaskan native. From this standpoint, it makes little difference whether such divergent abilities as have been found are native or acquired, if the presence of these abilities,

in quantities measurable by American measuring devices, is essential to success in the school system as at present organized. When we know beyond reasonable doubt that Eskimo or Aleut children actually measure appreciably lower than American children when measured with the same instruments in each of the several different fundamental abilities, very valuable basic information is at hand to use in planning curriculum, methods of instruction, and preparation of teachers which will assure a school system better adapted to their special abilities and instruction" (3, pp. 339-345).

Presumably, the Alaskan system of education will be improved by the prescription of a school system suited to American children with a mean IQ of 85. The implication is that the native ability of all races must be measured by an instrument based upon the average performance of American children and that the educational opportunities of such races should be based upon the results thereby obtained.

This is to neglect the obvious fact that intelligence can work through different cultural elements; that an American child who solves the "ball and field" test may be no more nor less intelligent than an Alaskan child who makes a flipper-toggle or a Samoan who makes a basket. One might hazard a guess that a mental test constructed and standardized by an intelligent Alaskan—and there are these even when measured by the Stanford-Binet scale (3, p. 345)—and applied to American children might result in an inordinate number of morons. This would be bad enough, but to reorganize our school system on the basis of such results would be folly indeed.

One might regret that such a program of testing was made the basis for racial comparisons. The results obtained are undoubtedly of value as measures within the racial groups, but neither the tests nor the techniques can warrant conclusions with respect to the comparison of Alaskan natives with the peoples of other cultures.

HAWAIIAN ISLANDS

The cosmopolitan nature of the population of Hawaii has invited a number of investigations in the field of race measurements. The pure Hawaiian population has dwindled, for various reasons, from approximately 300,000 in 1800 to 22,636 in 1930, when it constituted 6.1% of the total population (46). About half the total population is Asiatic; other large groups are Caucasian, part-Hawaiian, Portuguese, and Puerto Rican.

Livesay and Louttit (49) in 1930 made a study of differences in visual, auditory, and visual-choice reaction times, using 286 college

students. The differences were small and barely significant in favor of the Caucasian group. The coefficients of correlation between reaction times and intelligence were too low to be significant.

Louttit (52) in 1931 applied Porteus Maze, Healy Picture Completion, and Binet (Porteus-Babcock) tests to 224 boys and 137 girls (all part-Hawaiian) from the secondary department of the Kamehameha schools. The means on all tests increased with age up to 14 years and then remained fairly constant. For the boys, the mean IQ's on the Porteus Maze and the Binet (P-B) were around 100, and slightly less for the girls, probably due to some undetermined selective factors. Comparisons of groups on the basis of blood-mixture showed a superiority in favor of Hawaiian-White and Hawaiian-White-Chinese. The numbers tested were small, however, and the author suggests selective factors operative in the school which may have accounted for these differences. The differences for the blood-mixture groups were greater on the Binet (P-B) than on the other tests indicating, possibly, that the aim of the Porteus-Babcock-Binet "to eliminate tests which are immediately dependent upon language and those which require an oral response in well-phrased English" has not been completely realized.

Livesay (47) made comparisons in 1936 of 832 university students by the use of the American Council on Education Psychological Test. He found a significant difference in the total scores in favor of the Caucasian group. In an analysis of the scores in the separate sections of the test he found reliable differences in favor of the Caucasian group with respect to Completions, Analogies, Opposites, and Total Score. Chinese were superior in Artificial Language, while part-Hawaiians were inferior to all races except in the Completions. In a later analysis of the same group, Livesay (48) found significant differences in favor of males in Arithmetic and Analogies and of females in Artificial Language. In the Total Score there was a barely significant sex difference, the value of D/σ_D being 1.42. Livesay stresses the point that the results apply only to the groups measured and suggests that they are not typical of the parent groups because of the operation of selective immigration.

By the construction of the International Performance Scale, Leiter (44) hoped that he would be able to minimize the influence of cultural variables by the selection of common and familiar materials. His test consists of matching devices based upon Picture Completions, Color Forms, Associations, and the like. His scale was standardized upon 1430 Japanese and Chinese children in Hawaii, between the

ages of 6 and 17 years. The reliability of the test over the whole group by the split-halves method was .91.

In the age groups from 6 to 15 years the mean IQ's ranged from 98.4 to 102.6, the mean being approximately 100. Correlations between the Leiter scale and the Stanford-Binet and Porteus Maze were .79 and .71, respectively. The mean IQ for the Japanese group was 101.25 with a σ of 15.02, for the Chinese group 96.50 with a σ of 14.56. An application of the test to 146 children of Chinese-Hawaiian stock gave a mean IQ of 98, and application to 98 Caucasian children at the 6-, 9-, and 12-year levels gave mean IQ's of 114, 115, and 98, respectively. We do not know the basis of selection of the Caucasian group, and refrain from making comparisons.

It would be appropriate at this point to mention the application of the Leiter International Performance Scale to African natives by Porteus in 1934 (44, 80). The briefer scale was administered to 197 natives of different tribes, all individuals being above 15 years of age.

"To a people as primitive as these (Bushmen), even such a simple scale as the Leiter could not be properly applied. Without verbal instructions it seemed impossible to make the Bushmen understand what was required of them in the tests; this, in spite of considerable demonstration. . . . The tests were given to the Bakalakhadi . . . a somewhat degenerate tribe of Bechuana . . . without any spoken directions, so that the language difficulty did not appear to be a sufficient explanation of the Bushmen failure" (44, p. 25).

The mean mental age of the African group was about 10 years, as measured by the Leiter scale, which, it will be recalled, was standardized in Hawaii. Between Mission and "raw" natives there was a difference of little more than one year mental age in favor of the former. The numbers measured were small, and in view of the method of selection can hardly be regarded as typical. The selection was made by "confining the choice of subjects to the middle standards (so that) an average sample would be obtained (since) both the bright and dull groups would be eliminated" (44, p. 30). The school system is indeed fortunate that has succeeded in eliminating bright and dull children from its middle grades.

From the Hawaiian evidence we may infer that the Leiter scale is a useful device for the measurement of Chinese and Japanese children in the Hawaiian culture. Its international value is doubtful. The Leiter scale is apparently inapplicable to at least one African tribe. Unless we can be sure that the cultural variables have been held constant we should hesitate to hold the opinion that the perform-

ance of African natives is "at about the eleven-year level of performance as compared with that of Oriental children in Hawaii" (44, p. 33).

One is not sure that the Bushmen have had fair treatment. For instance, to administer in pantomime or in broken English to an immature person in America a test consisting of items almost wholly outside his realm of experience, such as abstract geometrical forms, strange animals, and peoples, and from its failure imply feeble-mindedness is psychometric moonshine; so too, to imply stupidity from the results of tests administered in a culture in which they are not applicable is unjustifiable.

In an investigation of Hawaiian-born children of Japanese, Chinese, part-Hawaiian, Filipino, and Portuguese descent, Porteus (80) found a fairly consistent Japanese superiority in five performance tests. The results of the Porteus test are shown in Table VI.

TABLE VI
MAZE TEST RESULTS IN HAWAII ON UNSELECTED BOYS

All Ages			14-Year Groups		
Race	N	Mean TQ *	Race	N	Mean TQ *
Japanese	228	102.0	Japanese	42	95.0
Part-Hawaiians	95	100.0	Part-Hawaiians	33	93.3
Filipinos	140	96.0	Chinese	36	92.0
Chinese	200	95.3	Filipinos	23	89.0
Portuguese	97	91.5	Portuguese	34	88.5

* Test Quotient.

It will be noticed that the differences are greater for the 'all ages' group than for the '14-year group.' It is quite possible that the selection of a relatively large sample of children of any one age will yield a more typical sample than the selection of an equal number of children of all ages. We are not given the sampling errors; therefore, the comparisons are difficult to make. Porteus is inclined to dismiss the sampling errors as unimportant.

"It will be noted that I have not calculated the significance of the differences between the Japanese and other races by using the formula usually applied in such cases. By means of this procedure the investigator might be able to state that the chances are 9,567 (or some such figure) in 10,000 that the Japanese are superior. Such a statement is merely ridiculous. We do not require one man to outrun another 9,000 times before we decide that he excels in running ability. The fact that the Japanese are superior in five trials and that the superiority is observable at each age level is quite sufficient" (80, p. 226).

Porteus admits that the "samples tested cannot be presumed to be wholly representative of the two races in their own countries. It can be assumed, however, that the two samples (Japanese and Chinese) were drawn from approximately equivalent levels in the two populations" (80, p. 225). With respect to the Portuguese, he suggests that "the tests used are only very partial measures of intelligence, in the broad sense that we have defined the term. It may well be that the tests used do not examine the special abilities characteristic of the Portuguese and hence do not represent their level of intelligence fairly" (80, p. 225).

In other words, we are in doubt, first, as to the representativeness of the population measured and, second, as to the reliability and validity of the tests used. Whatever comparisons we make must be tempered in the light of these two important reservations.

FIJI ISLANDS

The Colony of Fiji, situated in the south Tropical zone, consists of about 250 islands, 80 of which are inhabited. The population numbers 198,379 (Census, 1936), of which 97,651 are full-blooded Fijians, 85,002 are East Indians, 4028 are Europeans (from United Kingdom, Australia, and New Zealand), 4574 are mixed European and Fijian, and the remainder other Pacific Islanders and a small number of Asiatics.

Although the Fijians have had free cultural contact with Europeans since the middle of the Nineteenth Century, they have retained a great deal of their original culture and, except in the larger towns, still practice their communal system. Indians were introduced in 1879 to supply labor for the sugar-cane plantations, and of the total population of Indians more than 70% are Fijian-born. The Fijian is quite familiar with the school situation as illustrated by the fact that 85% are able to read and write Fijian, while a considerable number read and write English. Education has been a slower growth for the Indians, but there are signs of a developing interest in this direction.

In 1935 Mann (55, 56, 58) applied the Fiji Test of General Ability to more than 4000 Fijian and Indian children of school age in Fiji. After preliminary investigations in Fiji in 1934 Mann returned to Fiji and spent the greater part of 1935 in constructing and standardizing the Fiji Test of General Ability (55). It was felt that there was little to be gained by importing into Fiji a test standardized in another culture, and, since the validity of a test translated into Fijian and the nine or ten Indian vernaculars would be open to

question, a picture performance test was constructed, consisting of elements common to Fijian culture but independent of language. In the construction of the test the author had the opportunity of gaining some knowledge of Fijian and Indian cultural conditions, the assistance of Europeans familiar with cultural conditions, and the advice of Fijians and Indians of good intelligence and education. Preliminary forms were applied, and from the results a final form was constructed consisting of Classifications, Completions, Similars, Opposites, Analogies, Number Series, and Substitutions. This was applied to 2487 Fijians, 1169 Indians, 253 Europeans, and 177 half-castes drawn from all provinces and from every type of school.

Mann reports a correlation of .77 between the estimates of four reliable teachers and the Fiji test. The Goodenough test (drawing a man) was administered to 750 Fijian children and a correlation of .78 obtained between the results of the Fiji test and the Goodenough. The Otis Self-Administering Test (Intermediate Form) was applied to 250 European children, and a correlation between Otis and Fiji tests yielded a value of .74. The reliability of the test over ages 12 to 15 by the split-halves method was found to be .85.

Between the ages of 10 and 14 years the differences between the mean scores in the age levels are statistically significant. For Europeans at ages above 12 years the differences in mean scores at different age levels are not significant.

Although there are statistically significant differences between the scores of the race groups at the 12-, 13-, and 14-year levels, one must be wary of implying race differences from these results. The author insists that any value the Fiji test may have lies in its usefulness in classifying individuals within each race group. Following are some of his reasons for this attitude:

(a) The European population in Fiji is not typical of the parent population, consisting as it does of administrators, commercial executives, teachers, and missionaries. In 1934 the author found that the results of the Europeans in Fiji on a test of scholastic achievement standardized on 39,000 children in Australia were far above the Australian norms (54).

(b) In spite of the fact that age records are fairly reliable in Fiji, one cannot place too much faith in those used in the Fiji test. It would have been a very onerous task to have checked every one from birth records, and this was not done. In view of this, the scale scores on the test were used as age-norms, and no attempt was made to calculate mental ages or intelligence quotients.

(c) The infantile mortality rate for Fijians (1931-1935) was 112 per 1000, for Indians 73 per 1000, as compared with 63 for the United States, 47 for Australia, and 36 for New Zealand. The author was not able to discover whether the feeble-minded were deliberately neglected as infants and allowed to die, nor does he impute that this is done. It is significant, however, that in all of the many schools visited no children were found with obvious mental defect. The typicality of the samples used must remain in doubt awaiting more evidence on this point.

(d) The children tested were all school children. The report of the school population indicates that while 65% of Fijian children are in school, no more than 26% of Indians are enrolled, and of these, for social and religious reasons, less than one-third are girls. It might be assumed that the Fijian sample is fairly representative of the Fijian population, but the Indian sample is definitely atypical.

(e) The percentage of overlap in the Fijian and Indian races leads one to express comparisons of these races with caution. In the 12-year group, although the differences in mean scores of Fijians and Indians are significant—the value of D/σ_D being 3.25—it is likely that 40% of the Fijians are equal or superior to the mean of the Indians.

Mann also applied the Seashore Test of Musical Talent to more than 800 children of Fijian and Indian races. In view of the obvious difficulty the children experienced in understanding and carrying out the directions, the results were discarded.

The conclusion arrived at from the results of this testing program is that there is no evidence of any value relative to a valid comparison of these races (58).

AUSTRALIA

The aboriginal population of Australia in 1936 (Census, 1936) consisted of 53,698 fullblood aborigines and 23,461 half-castes, making a total of 77,159. Because of their nomadic habits it has been found difficult to test them in large numbers. Porteus (80) sums up the recent programs which have used the Porteus Maze Test.

These small groups, making a total of little more than 200, were from various tribes scattered over an area as large as the United States. Most of them were from mission stations. In view of the number tested—200 in more than 50,000—and of the factors operative in the selection of natives for the missions, conclusions from these data would need some qualification. Porteus believes that "the

average test age of the five groups of children examined (by him) would be about 10.5 years. Chronologically they would average about 12½ years, which would make their average test quotient about 84" (80, p. 240).

TABLE VII

MEAN MENTAL AGES OF AUSTRALIAN ABORIGINES AS MEASURED BY THE PORTEUS MAZE TEST

Subjects	Examiner	N	Mean Mental Age
Half-caste boys	Stoneman	13	11.1
Half-caste girls	Stoneman	20	10.6
Boys over 10 years	Piddington	—	10.5
Girls over 10 years	Piddington	—	10.1
Children mixed	Porteus	22	10.1
Adult females	Piddington	14	8.6
Adult females	Porteus	11	10.1
Adult males	Porteus	25	12.1
Adult males	Porteus	14	11.3
Adult males	Fry-Pulleine	10	10.7
Adult males	Piddington	24	10.5
Adult males	Porteus	65	10.5

AFRICA

In recent years there have been attempts to apply suitable tests of general ability to natives of Africa. In 1933, after experimental work with several well-known tests, Oliver developed the General Intelligence Test for Africans (68). This was a picture test which could be administered in any language to native children of East Africa. The test was standardized on several hundred children. Over 100 boys in Grades V and VI and 67 boys in Grades I and II it yielded a validity of .6 and a reliability of .8. The work of establishing norms was not completed. The test seemed to prove satisfactory as a measure of the ability of the natives to whom it was applied. In view of the difficulty of controlling the culture variables, Oliver refrained from making race comparisons (67, 68, 70, 71, 72).

In 1934 Porteus conducted an expedition to parts of South Africa and there administered performance tests to groups of natives of different tribes (80). He believes that the problem of races is not a question of superiority but of differences. However, in view of the fact that most of his investigations are in the direction of quantification, it is obvious that any differences he claims will inevitably lead to a belief in superiority, at least in the trait measured. He writes:

"All the studies that the writer has made into the question dispose him towards the belief in the cumulative effects of environment in determining

the character of peoples. Climatic and other physical conditions have a selective effect, and, if continued long enough, seem to make an indelible impress on the physical and mental constitution of the inhabitants of a country. . . . Hence it is easy to suppose that natio-racial differences, though related to environment, ultimately through selection become biological" (76, pp. 183-184).

This is a contention that is rather easier to suppose than it is to prove. More recently he has argued:

"While the physical differences are being set by heredity, the same differentiation from other races may be brought about with regard to mental factors. Hence, one nascent race may vary from another on the average as regards both general and special mental abilities" (80, p. 5).

Confining his comparisons to the results obtained from Australian and African natives, Porteus is inclined towards the belief that he has evidence for the superiority of the Australian native.

TABLE VIII

RESULTS OF MAZE TEST PERFORMANCE BY AUSTRALIAN AND AFRICAN NATIVE PEOPLES (Porteus, 80, p. 257)

Tribe	Locality	Schooling	N	Mean Test Age	σ
Arunta	C. Australia	Mission	25	12.08	2.09
Bathonga	N. Transvaal	Mission	29	11.72	2.22
Wakaranga	S. Rhodesia	Mission	32	11.57	2.17
Ndau	S. Rhodesia	Mission	43	11.41	2.20
Mixed	W. Australia	Government School	14	11.32	—
Amaxosa	Cape Province	None	25	10.78	2.76
Karadjeri	N. W. Australia	None	24	10.52	2.60
Keidja-Nyul	N. W. Australia	Mission	65	10.48	2.34
Shangans	Port E. Africa	None	25	9.30	2.66
Mchopi	Port E. Africa	None	28	8.34	2.45
Bushmen	Kalahari	None	25	7.56	2.17

If the results are redistributed, however, it will be seen that the mean Test Age for the 128 Australian natives is 10.89 and for the 207 African natives, 10.27. Using a σ of 3.0, the value of D/σ_D is 1.8. For the 208 mission-schooled natives—both Australian and African—the mean Test Age is 11.26, and for the 127 unschooled natives of both races, 9.27. Using a σ of 3.0, the value of D/σ_D is 6.0. It is apparent, then, that the significance of the difference between schooled and unschooled natives is much greater than that between race groups.

Porteus also submits the results of tests of brain capacity, hand grip (strength and dominance), Goddard formboard, auditory and visual rote memory, Form and Assembly Test, Footprints Test, and the Leiter International Performance Test. In all cases the numbers

were small, and differences of varying degrees of significance were reported.

If the results submitted are to be regarded as evidence for innate differences, it would appear necessary to accept at least two assumptions. In the first place, we would have to assume that the small groups tested (averaging few more than 30 in each tribe tested) were truly representative of the tribes from which they were drawn. In view of the environmental difficulties and of the nomadic nature of the groups, this is unlikely. Porteus indicates this with respect to the Bushmen (80, p. 241), regretting that he did not visit areas where they were more numerous—and perhaps more typical. The difficulties of securing random samples may have operated with other tribes.

In the second place, the assumption would have to be made that the Porteus Maze and other tests were equally applicable to African and Australian natives, Asiatics in Honolulu, and defectives in Vineland. In other words, they are universal tests; although standardized in but one culture, they are equally applicable to any other, measuring equally well in every culture some simple or complex trait. This is a claim that is made for no other test. What we have are comparisons made on the results of a test which in one culture is claimed to measure a "complex of qualities . . . (which) seems to be valuable in making adjustments to *our* kind of society," applied to small, but not necessarily typical, groups in other cultures. It might quite well be that it is the test (and not the subjects) which is being tested.

Finally, even if the above assumptions be accepted, and there will be few who are willing to accept them without qualification, the most significant differences in mean Test Scores are not between racial groups, but between groups which have and which have not had such educational opportunities as are offered primitive peoples. The results likely to be obtained with more adequate opportunities for schooling are open to conjecture.

CONCLUSIONS

During the past decade attempts at the measurement of primitive races have followed three main trends:

- (1) Investigations of the physical and psychophysical traits of primitive peoples.
- (2) Construction and standardization of tests within a native

culture for the purpose of classifying individuals within the culture of the test.

(3) Application of tests standardized in one culture to native groups of other cultures.

The difficulties in the way of race comparisons have already been reviewed. The criteria to be met are (1) the availability of tests which are unequivocally valid in the culture in which they are used, and (2) the selection of cases which are truly representative of the racial groups under review. A suggestion by Smith (87) to incorporate factor analysis is promising, but, as far as is known, has not yet been acted upon.

For the present, however, it must be admitted that the evidence assembled for primitive peoples has not met these criteria of comparison, and until it does, or until new and valid techniques are established, the problem of race differences among primitive peoples remains unsolved.

BIBLIOGRAPHY

1. ANASTASI, A. *Differential psychology*. New York: Macmillan, 1937.
2. ANASTASI, A., & FOLEY, J. P., JR. An analysis of spontaneous drawings by children of different countries. *J. appl. Psychol.*, 1936, 20, 689-726.
3. ANDERSON, H. D., & ELLS, W. C. *Alaska natives: a survey of their sociological and educational status*. Stanford Univ.: Stanford Univ. Press, 1935.
4. ARLITT, A. H. On the need for caution in establishing race norms. *J. appl. Psychol.*, 1921, 5, 179-183.
5. BACHE, R. M. Reaction time with reference to race. *Psychol. Rev.*, 1895, 2, 474-486.
6. BELL, R. *Intelligence*. In Strong, E. K., Jr., *Vocational Aptitudes of Second-Generation Japanese in the United States*. Stanford Univ.: Stanford Univ. Publ. Educ.-Psychol., 1933, Ser. 1, No. 1.
7. BELL, R. Public school education of second-generation Japanese in California. *Stanford Univ. Publ. Educ.-Psychol.*, 1935, Ser. 1, No. 3.
8. BENEDICT, R. *Patterns of culture*. New York: Houghton Mifflin, 1934.
9. BLACKWOOD, B. A study of mental testing in relation to anthropology. *Ment. Meas. Monogr.*, 1929, No. 4.
10. BOAZ, F. *The mind of primitive man*. New York: Macmillan, 1938.
11. BOYD, W. C., & BOYD, L. G. Sexual and racial variations in ability to taste phenyl-thio-carbamide with some data on inheritance. *Ann. Eugen., Camb.*, 1937, 8, 45-51.
12. BRACE, D. K. *Measuring motor ability*. New York: Barnes, 1927.
13. BRUNER, F. G. The hearing of primitive peoples. *Arch. Psychol., N. Y.*, 1908, 2, No. 11.
14. BRUNER, F. G. Racial differences. *Psychol. Bull.*, 1914, 11, 483-486.

15. CIPRIANI, L. [Race and mentality apropos of miscegenation with Africans.] *Rass. int. Clin. Terap.*, 1936, 17, 584-590.
16. DANIEL, R. P. Basic considerations for valid interpretations of experimental studies pertaining to race differences. *J. educ. Psychol.*, 1932, 23, 15-27.
17. DARSIE, M. The mental capacity of American-born Japanese children. *Comp. Psychol. Monogr.*, 1926, 3, 15.
18. DENSMORE, F. The native music of American Samoa. *Amer. Anthropol.*, 1932, 34, 415-417.
19. DUNLAP, J. W. Race differences in the organization of numerical and verbal abilities. *Arch. Psychol., N. Y.*, 1931, 19, No. 124.
20. ELLS, W. C. Mental ability of the native races of Alaska. *J. appl. Psychol.*, 1933, 17, 417-438.
21. ELLS, W. C. Mechanical, physical and musical ability of the native races of Alaska. *J. appl. Psychol.*, 1933, 17, 493-506.
22. ELLS, W. C. Educational achievement of the native races of Alaska. *J. appl. Psychol.*, 1933, 17, 646-670.
23. ESTABROOK, G. H. A proposed technique for the investigation of racial differences in intelligence. *Amer. Nat.*, 1928, 62, 78-87.
24. FICK, M. L. A mental survey of the Union of South Africa. *S. Afr. J. Psychol. Educ.*, 1932, 1, 31-46.
25. FREEMAN, F. N. The interpretation of test results with especial reference to race comparisons. *J. Negro Educ.*, 1934, 3, 519-522.
26. FRY, H. K., & PULLEINE, R. H. The mentality of the Australian aborigine. *Aust. J. exp. Biol. med. Sci.*, 1931, 8.
27. GARTH, T. R. A review of race psychology. *Psychol. Bull.*, 1930, 27, 329-356.
28. GARTH, T. R. Race psychology. A study of mental differences. New York: McGraw-Hill, 1931.
29. GARTH, T. R. The problem of race psychology: a general statement. *J. Negro Educ.*, 1934, 3, 319-327.
30. GILLILAND, A. R., & CLARK, E. L. Psychology of individual differences. New York: Prentice-Hall, 1939.
31. GOLDENWEISER, A. Anthropology. New York: Crofts, 1938.
32. GOODENOUGH, F. L. Measurement of intelligence by drawings. Yonkers, N. Y.: World Book, 1926.
33. GOODENOUGH, F. L. The measurement of mental functions in primitive groups. *Amer. Anthropol.*, 1936, 38, 1-11.
34. HARMON, C. Racial differences in reaction time at the preschool level. *Child Devolpm.*, 1937, 8, 279-281.
35. HARRASSER, A. Konstitution und Rasse. 1933, 1934, 1935, 1936. *Fortschr. Neur. Psychiat.*, 1937, 9, 471-490.
36. HOOTON, E. A. Up from the ape. New York: Macmillan, 1938.
37. HOUSE, F. N. Viewpoints and methods in the study of race relations. *Amer. J. Sociol.*, 1935, 40, 440-452.
38. JULIEN, P. F. J. A. [The distribution of the taste threshold for phenylthio-urea in the Netherlands and in the western equatorial Africa.] *Mensch en Maatsch.*, 1938, 14, 364-365.

39. KLINEBERG, O. Racial differences in speed and accuracy. *J. abnorm. soc. Psychol.*, 1927, 22, 273-277.
40. KLINEBERG, O. Cultural factors in intelligence-test performance. *J. Negro Educ.*, 1934, 3, 478-483.
41. KLINEBERG, O. Race differences. New York: Harper, 1935.
42. KROEGER, A. L. Anthropology. New York: Harcourt, Brace, 1923.
43. LEHMAN, H. C., & WITTY, P. A. Racial differences: the dogma of superiority. *J. soc. Psychol.*, 1930, 1, 394-418.
44. LEITER, R. G. The Leiter International Performance Scale. (With an appendix by S. D. Porteus.) *Univ. Hawaii Res. Publ.*, 1936, Ser. 13, No. 15.
45. LINTON, R. The study of man. New York: Appleton-Century, 1936.
46. LIVESAY, T. M. A study of public education in Hawaii. *Univ. Hawaii Res. Publ.*, 1932, 7.
47. LIVESAY, T. M. Racial comparisons in performance on the American Council psychological examination. *J. educ. Psychol.*, 1936, 27, 631-634.
48. LIVESAY, T. M. Sex differences in performance on the American Council psychological examination. *J. educ. Psychol.*, 1937, 28, 694-702.
49. LIVESAY, T. M., & LOUTTIT, C. M. Reaction time experiments and racial groups. *J. appl. Psychol.*, 1930, 14, 557-565.
50. LONG, H. H. On mental tests and racial psychology: a critique. *Opportunity*, 1925, 3, 134.
51. LOUTTIT, C. M. Racial comparison of ability in immediate recall of logical and non-sense material. *J. soc. Psychol.*, 1931, 2, 205-215.
52. LOUTTIT, C. M. Test performance of a selected group of part-Hawaiians. *J. appl. Psychol.*, 1931, 15, 43-52.
53. MACQUARRIE, T. W. A mechanical ability test. *J. Person. Res.*, 1927, 5, 329-337.
54. MANN, C. W. Education in Fiji. Melbourne: Melbourne & Oxford Univ. Press, 1935. (*Aust. Coun. educ. Res. Ser. 33.*)
55. MANN, C. W. Fiji Test of General Ability (Handbook). Suva, Fiji: Government Printer, 1935.
56. MANN, C. W. Objective tests in Fiji. Suva, Fiji: Government Printer, 1937.
57. MANN, C. W. The educational system of the Colony of Fiji. Unpublished Doctoral Dissertation, Stanford Univ., 1937.
58. MANN, C. W. A test of general ability in Fiji. *J. genet. Psychol.*, 1939, 54, 435-454.
59. MEAD, M. An investigation of the thought of primitive children, with special reference to animism. *J. R. anthropol. Inst.*, 1932, 62, 173-190.
60. MEAD, M. The primitive child. In Murchison, C. (Ed.), *Handbook of Child Psychology*. Worcester, Mass.: Clark Univ. Press, 1933. Pp. 909-927.
61. MEAD, M. The use of primitive material in the study of personality. *Character & Pers.*, 1934, 3, 1-16.
62. MEAD, M. Sex and temperament in three primitive societies. New York: Morrow, 1935.

63. MEAD, M. The methodology of racial testing: its significance for sociology. *Amer. J. Sociol.*, 1936, 21, 657-667.
64. MERRY, R. C. Art talent and racial background. *J. educ. Psychol.*, 1938, 32, 17-32.
65. NISSEN, H. W., MACHOVER, S., & KINDER, E. F. A study of performance tests given to a group of native African children. *Brit. J. Psychol.*, 1935, 25, 308-355.
66. OBERLY, H. S. Preliminary report of experiments with West African Negroes. *Psychol. Bull.*, 1935, 32, 558-559.
67. OLIVER, R. A. C. The comparison of abilities of races: with special reference to East Africa. *E. Afr. med. J.*, 1932 (September), 160-204.
68. OLIVER, R. A. C. General intelligence test for Africans (with manual of directions). Nairobi, Kenya Colony: Government Printer, 1932.
69. OLIVER, R. A. C. The musical talent of natives of East Africa. *Brit. J. Psychol.*, 1932, 22, 333-343.
70. OLIVER, R. A. C. The adaptation of intelligence tests to tropical Africa. *Overseas Educ.*, 1933, 4, 186-191.
71. OLIVER, R. A. C. The adaptation of intelligence tests to tropical Africa. II. *Overseas Educ.*, 1933, 5, 8-13.
72. OLIVER, R. A. C. Mental tests in the study of the African. *Africa*, 1934, 7, 40-46.
73. PECK, L., & HODGES, A. B. A study of racial differences in eidetic imagery in preschool children. *J. genet. Psychol.*, 1937, 51, 146-161.
74. PETERSON, J. Basic considerations of methodology in race testing. *J. Negro Educ.*, 1934, 3, 403-410.
75. PINTNER, R. The influence of language background on intelligence tests. *J. soc. Psychol.*, 1932, 3, 235-240.
76. PORTEUS, S. D. Race and social differences in performance tests. *Genet. Psychol. Monogr.*, 1930, 8, 83-208.
77. PORTEUS, S. D. The psychology of a primitive people. New York: Longmans, Green, 1931.
78. PORTEUS, S. D. Human studies in Hawaii. Pacific problems. *Proc. Sch. Orient. Pacif. Affairs, Univ. Hawaii*, 1932, 82-114.
79. PORTEUS, S. D. The Maze Test and mental differences. Vineland, N. J.: Smith Printing & Publishing House, 1933.
80. PORTEUS, S. D. Primitive intelligence and environment. New York: Macmillan, 1937.
81. PRIEST, J. Bible defense of slavery; and origins, fortunes and history of the Negro race. Glasgow, Ky.: Brown, 1853.
82. RIVERS, W. H. R. Observations on the senses of the Todas. *Brit. J. Psychol.*, 1904, 1, 452-468.
83. SANDERSON, H. E. Differences in musical ability in children of different national and racial origin. *J. genet. Psychol.*, 1933, 42, 100-119.
84. SAYCE, R. U. Primitive man and civilised man. *Scientia, Milano*, 1935, 57, 53-62.
85. SEASHORE, C. E. Measures of musical talent. Chicago: Stoelting, 1919.
86. SELTZER, C. C. A critique of the coefficient of racial likeness. *Amer. J. phys. Anthropol.*, 1937, 23, 101-109.

87. SMITH, C. E. A new approach to the problem of racial differences. *J. Negro Educ.*, 1934, 3, 523-529.
88. STEGGERDA, M. Racial psychometry. *Eugen. News*, 1934, 19, 132-133.
89. STEGGERDA, M. Testing races for the threshold of taste with PTC. *J. Hered.*, 1937, 28, 309-310.
90. STRONG, E. K., JR. Vocational aptitudes of second-generation Japanese in the United States. *Stanford Univ. Publ. Educ.-Psychol.*, 1933, Ser. 1, No. 1.
91. STRONG, E. K., JR. Japanese in California. *Stanford Univ. Publ. Educ.-Psychol.*, 1933, Ser. 1, No. 2.
92. THOMPSON, C. H. The conclusions of scientists relative to racial differences. *J. Negro Educ.*, 1934, 3, 494-512.
93. THOULESS, R. H. A racial difference in perception. *J. soc. Psychol.*, 1933, 4, 330-339.
94. WALTERS, F. C. Language handicap and the Stanford Revision of the Binet-Simon tests. *J. educ. Psychol.*, 1924, 15, 276-284.
95. WILKERSON, D. Racial differences in scholastic achievement. *J. Negro Educ.*, 1934, 3, 453-477.
96. WOODWORTH, R. S. Racial differences in mental traits. *Science*, 1910, 31, 171-178.
97. WOODWORTH, R. S. Comparative psychology of the races. *Psychol. Bull.*, 1916, 13, 388-396.
98. YODER, D. Present status of the question of racial differences. *J. educ. Psychol.*, 1928, 19, 463-470.

BOOK REVIEWS

WHEELER, R. H. *The science of psychology: an introductory study.* (2nd rev. ed.) New York: Crowell, 1940. Pp. xviii+436.

New general textbooks of psychology must be evaluated in terms of their major aim. Some are mainly *system-making* books, whose aim is to present a new point of view, to sell the reader on this viewpoint, and to show how the subject matter of psychology can be effectively organized about it. An example is Watson's *Psychology from the standpoint of a behaviorist*. Such books are only secondarily teaching texts. They are mainly polemical. Others have as a major purpose some innovation in organization of topics and subject matter, with the idea of presenting the science more effectively to the student, making it more "practical," or boiling it down to essentials. Considerable originality can be displayed in either of these types of text, and they may, in the long run, exert a large influence in molding the future development of the subject. But at least initially, they present a problem to the average teacher. To use them means to burn away old barriers, to adopt new thought grooves. The student has no such negative transfer effect to overcome. But student and teacher have to talk the same language. Another source of hesitancy in adopting a radically new text is the natural skepticism about untried things until the scientific fraternity has given them the stamp of approval.

The present text is a decided innovation on both the above counts. It is, first, a system-making book. It aims to present the system of 'organismic' psychology. Since this term has already been used by other system-makers with a different meaning, it is necessary to give a more precise characterization of the Wheeler system. Several philosophers, e.g. Whitehead, have called attention to the 'organismic' character of nature, of the universe as a whole, implying that no part behaves independently, but the behavior of each is dependent on the principles of organization of the whole. The analogy with the living organism, and its self-maintained integrity, is obvious. To this core Wheeler has added the doctrine of the Emergentists that the properties of new complex groupings of elements do not pre-exist in the elements, but come into being at the time of their union. Finally, he has followed a parallel course with Gestalt psychology in attributing determinative power to configurations, as such, in the development of new integrations, mental or behavioral, like problem-solving, perception, etc. He has adapted some of the concepts of topological psychology, such as 'barriers,' and has received from this school a strong slant toward social-psychological interpretations. Wheeler's system is not to be identified, however, with any one of the above-mentioned schools. He has formulated many laws of his own, which are better described as interpretative principles, such as the Law of Determined Activity, and has adapted some from physical science,

making an unorthodox use of them, as the Law of Action and Reaction. He believes that the time has come when psychology should formulate laws and that this should precede exact measurement. Laws are, to him, statements of general direction or tendency rather than precise inductive formulations. They are intuitive insights from meager data, yet somehow possess the authoritativeness of revelations.

Secondly, Wheeler's text is an innovation in the other of our two senses, in that he has projected a new order of development of topics to fit his psychological viewpoint. If psychology is organismic, then all principles and processes derive their significance from the larger whole. For psychology, the largest whole is the social organism. Individual behavior derives its meaning from its function in the social organism, primarily. Hence, the first topic to be dealt with, after the preliminary orientation, is Social Behavior and Its Conditions. Then, by a process of greater and greater 'individuation,' come Development and Measurement of Personality, Emotive Behavior, Intelligent Behavior, Learning, Observational Behavior, including sensory and perceptual processes, and, last, the Nervous System. Not only does the logic of whole-to-part organization apply to the order of topics, but each topic is similarly approached. For example, Personality is first treated genetically as an integrated thing and then from the point of view of measurement of separate traits.

With this brief survey before us, we can attempt a rough evaluation of the book. First, what of the system of 'organismic' psychology? I believe few present-day psychological thinkers would deny that there has been an overemphasis in the past on the logic of the analytical approach to human behavior. The revolt which Wheeler is championing is not new. Perhaps he is the first to make it the major keynote of an introductory text. In so doing, he has weakened the book as a text. The constant reiteration of the theme that parts derive their meaning from wholes finally leads one to exclaim: "Methinks he doth protest too much." Student readers will accept the simple proposition quite readily. If he is striving to convince his psychological contemporaries, who are supposedly so steeped in atomistic mechanism that they must be converted, an introductory text is no place for this battleground.

But 'organismic' psychology is more than just a focus of emphasis. Wheeler has formulated a large number of "laws" and has interpreted the whole field of psychology in terms of them. A major part of the discussion of each topic is concerned with interpreting the phenomena as expressions of these general "laws." We can have no quarrel with such a procedure *per se*. It is an ideal way to organize a science text, provided the "laws" are well established and sound. On the other hand, to the extent that their soundness is questionable, the entire system is weakened. What are some of these laws? He presents, in the first chapter, the following: the Law of Field Properties, that "wholes exist in their own right over and above the parts or ingredients from which, through closure, they were formed"; the Law of Determined Activity, that "the whole regulates the activities of its parts"; the Law of Derived Properties, that

"the properties of the parts are derived from the wholes of which they are members"; the Law of Individuation, that "parts come into existence through a division process that can be called individuation." Note that these "laws" are completely general, so that wherever, for example, individuation in behavior occurs, it is explained by saying that "parts come into existence through a division process." To these are added twelve others in later chapters. Least Action and Closure are already familiar. The discussion of emotion calls for Laws of Action and Reaction and Maximum Work. Intelligence demands Laws of Transposition, Configuration, and Insight; Learning calls for Reciprocal Change, Permanence, and Increasing Energy; Perception requires also Laws of New Insight and Field Genesis. One feels that the author, like Jehovah in "Green Pastures," is ready to "r'ar back and pass a miracle" and create a new law for every emergency.

That these so-called laws are not precise or experimentally proved principles does not bother him because he holds that there are three stages in the refinement of science and that psychology is only now at that stage where general statements of direction or tendency are possible.

To the reviewer, these "laws" are no more than statements of broad analogy, or rough descriptive phrases. They need to be taken in a somewhat figurative sense, just as the statement that "element x has an affinity for element y" is figurative in chemistry. Others of them, like the Law of Permanence, are only tautological reiterations of the observed facts which they purport to explain. Is memory explained by appealing to a law of "permanence"? Quite a wide range of experimental material is discussed in the process of explaining and illustrating the laws, which helps to enrich the content of the book. For those who desire a clearer idea of the nature of 'organismic' psychology, this book will be of value. As a teaching text, it is likely to have a rather limited appeal.

ARTHUR G. BILLS.

University of Cincinnati.

WESTERHOF, A. C. Representative psychologists. Union Bridge, Md.: Pilot, 1938. Pp. vi+119.

This slender book rests on the thesis that nothing differentiates psychologists more characteristically than the way in which they deal with "the problem of mechanism and teleology" (p. v). In the light of this thesis the ideas of a considerable number of psychologists are examined. The criteria whereby these men are selected are not clear, though there is no question that, with the exception of one man (Pauly) who is not a psychologist, all are representative of aspects of contemporary psychology. There are others, however, who are equally representative.

There is a separate chapter on each man in which the implications of his writing for mechanism or teleology are stated and in which Westerhof gives his own critical evaluation of the man's ideas. The Persons of Critical Personalism are "teleological wholes" (p. 22), but Stern is said to err in placing disposition teleology above intention teleology. Koffka

and Köhler recognize that behavior is purposive, but, nevertheless, they make it seem too little different from physical process. Lewin rejects goal-seeking interpretations of behavior and regards goal-directedness as forced upon the organism by the object. Westerhof finds "a discernible circularity" (p. 53) in Lewin's discussion. The fact that Terman has refrained from overt theoretical construction does not deter Westerhof from asserting, on the basis of Terman's discussions of the individual tests in the Stanford Revision, that for him "conscious insight is a determining factor in behavior" (p. 110). One wonders if Terman means all that is implied in this chapter.

Tolman, for whom purpose and cognition are descriptive terms, tries to avoid "recognition of consciousness as significant for behavior" (p. 110), in spite of the fact that his data should have forced upon him the primacy of consciousness. Freud and Jung display a "teleology phobia accompanied by consciousness phobia" (p. 110), while Adler, who, of the three, comes nearest to being purposive, is not critically teleological. On the other hand, Pauly, the Lamarckian biological theorist, "assumes that consciousness and teleology extend down even into the inorganic realm" (p. 110), but his constructive theory is said to be weak. McDougall's argument for the causal efficacy of consciousness is powerful evidence for teleology, in spite of the difficulties which attend his concept of energy.

A summary and appendix present the argument *about*, and the argument *for*, teleology. We are told that the besetting sin of psychologists is their fixation on raw experimental data and their corresponding failure to realize that consciousness is always with us. Thinking, as an example of conscious process, is a "unique form of striving" (p. 118). The relation between consciousness and teleology is not carefully worked out, and a large amount of recent writing which is directly relevant to the problem of consciousness is given no mention. Westerhof admits that the argument for teleology rests on no single crucial datum. He insists, rather, that "the argument for teleology grows as the organization of our knowledge increases" (p. 113).

The book shows no recognition of the diverse meanings which mechanism and teleology may have. The first chapter reviews certain characteristics of each, but achieves no incisive analysis of what either means. Whether it is profitable now for psychology to be concerned with the problem will be decided by the individual reader's theoretical perspective. If one does wish to be concerned, this book provides an introduction to certain aspects of the problem, but by no means to all of them. The book's fundamental thesis that psychologists are characteristically differentiated by their handling of the mechanism-teleology problem may be doubted.

Westerhof sometimes cites older books or editions when later ones are available and are more complete statements of a man's thinking. There is no index.

JOHN A. McGEOCH.

University of Iowa.

UNIVERSITY OF IOWA LIBRARIES

LÖWENFELD, V. *The nature of creative activity: experimental and comparative studies of visual and non-visual sources of drawing, painting, and sculpture by means of the artistic products of weak sighted and blind subjects and of the art of different epochs and cultures.* New York: Harcourt, Brace, 1939. Pp. xvii+272.

The content of this book is indicated more correctly by its subtitle. Dr. Löwenfeld has not explained the nature of creative activity. He has, however, given us three things: a method, a mass of material, and a theory. The importance of each will be weighted in accordance with the bias of the reader, but there is no doubt that the book as a whole represents an important contribution both to psychology and to aesthetics. Dr. O. A. Oeser is to be complimented on an excellent job of translation.

The originality of Löwenfeld's method lies in his use of weak-sighted subjects. The psychological literature on artistic creation contains excellent studies of artistic productions of children, of the blind, and of primitive people, but the attempt to find principles of creativity common to all three has not been strikingly successful. There is undoubtedly a parallel between the drawings of children and of primitives, and the sculpture of both bears a striking resemblance, in some respects, to the sculpture of the blind. Exceptions to these parallels are so numerous, however, as to vitiate any simple theory based on the assumption of developmental stages. The difficulty, Löwenfeld believes, arises partly from the interpretation in purely visual terms of what may be essentially nonvisual work, partly from an unwarranted emphasis on completed products as distinguished from the process of production, and partly from a failure to discount the influence of such factors as inadequate motor coördination, which are essentially irrelevant to the creative process as such. The weak-sighted subject, like the normal child, makes use of visual material in his artistic work, but tends to be constrained by his defect to use methods characteristic of the blind. The results, Löwenfeld finds, permit the psychologist to make a clear distinction between visual and haptic types of perception and creation. The dominance of the haptic is what is characteristic of most of the artistic productions of the blind, and it is the emphasis on this component of experience which serves as the connecting link between primitive art and the art of childhood.

Löwenfeld bases his findings on the drawings, paintings, and sculpture of weak-sighted subjects, ranging in age from eight to twenty years, the work of fifteen of whom is reproduced in the book. For purposes of comparison he includes examples from the drawing of a few normal children and from the painting and sculpture of primitives. In all, there are more than 200 individual reproductions, ranging in theme from simple attempts to reproduce the human form to representations of such a complicated subject as: "A beggar goes over the street, is knocked down by a car, and loses hat and money." It is a fascinating collection and is well worth study, quite apart from the stimulating interpretation which accompanies it. The author begins with an analysis of the principal characteristics of children's drawings, emphasizing their expressionistic character and showing how this is related to the predominance in children of the

haptic type of perception. This forms the basis for a more systematic discussion of what he considers to be the two principal creative types, the visual and the haptic, a distinction which is further supported by a detailed analysis of the drawings of weak-sighted subjects and a comparison of these with the artistic productions of the blind. In a final chapter he applies his theory to the interpretation of primitive art and attempts to account in terms of his two types for the age-old conflict between impressionism and expressionism.

A theory which is couched in typological language is likely to be greeted with suspicion. A typology creates the appearance of an explanation when it has done nothing more than state a problem. But a well-formulated problem is in itself a contribution of importance—*prudens quaestio dimidium scientiae*—and it must not be forgotten that the very postulate of a typology implies the prior recognition of certain important directions of variation in the phenomena which are being studied. The value of Löwenfeld's work for psychology lies not in his designation of two new psychological types, although he has presented them without undue dogmatism, but in the fruitfulness of his phenomenological analysis. The phenomenological method in recent studies of perception has helped to restore the self (ego, person) to its proper place in the psychological field and to show how such elementary properties as distance, size, and weight vary in accordance with their degree of subjectivity (felt dependence on self) or objectivity (independence of self-reference). If subjectivity and objectivity are not only phenomenal, but also functional, properties of the psychological field, it follows that a person who lives in a highly "subjectified" world will differ in his methods of pictorial representation from a person whose world is more thoroughly "objectified." The subjective emphasis will lead to a more highly functionalized type of representation, with sizes and shapes distorted to express the affective tendencies of the artist. The world of the blind is such a subjectified world, and the art of the blind shows corresponding tendencies. In children's drawings the same emphasis is present, although frequently obscured by the technical incompetence of the child. It is in the artistic work of the weak-sighted, Löwenfeld finds, that the rival tendencies express themselves most clearly; for a person who is almost blind lives in a world which contains the fundamental visual components but which lacks the objective independence and stability of the world of the normal person.

To reduce the whole matter to a dichotomy of haptic and visual types seems to the reviewer to be somewhat unfair to the facts. The variables in the psychological field are too complex to be disposed of in terms of the traditional sense modalities, and, in any case, Löwenfeld has not demonstrated that his types are really sensory types. If the term "type" is justified at all, it should refer here to characteristics of field-structure which are more fundamental than any modal distinctions. Fortunately, however, Löwenfeld's basic observations are not affected by such criticisms. His central purpose was not to present a psychological theory, but to indicate some of the essential characteristics of the creative process—and this he has done admirably. In the reviewer's opinion, the book is

a valuable contribution in itself and contains a wealth of suggestions for productive experimentation.

R. B. MACLEOD.

Swarthmore College.

SCHEIDEMANN, N. V. *Lecture demonstrations for general psychology.* Chicago: Univ. Chicago Press, 1939. Pp. x+241.

SCHEIDEMANN, N. V. *Experiments in general psychology.* (Rev. ed.) Chicago: Univ. Chicago Press, 1939. Pp. xiv+201.

It is undoubtedly true that a well-planned demonstration enlivens a class, tends to arouse the students' interest, and may serve a useful pedagogical function. If carried to excess, demonstrations will interfere with the more serious aspects of instruction. The author of *Lecture demonstrations for general psychology* does not expect that an instructor will use in any single course all of the sixty demonstrations that are included in this book, which range, incidentally, all the way from sensory phenomena to "extra-sensory perception: clairvoyance and telepathy." The instructor can choose according to his interests and emphases in a course.

In the words of the author, the purpose was "to organize, adapt, and condense various reported investigations into simple and concrete demonstrations that may be performed in connection with lectures in general psychology. Each demonstration is based upon, and follows very closely, the experimental work of a recognized teacher of psychology. For each demonstration the purpose, the material required, the steps of procedure, and the points of interest to the class are stated definitely. Following each demonstration are summary comments on the original experimenter's findings and conclusions." It is the opinion of the reviewer that this purpose has been faithfully executed. The material needed for each experiment can easily be assembled by the instructor; no special equipment or apparatus is called for.

Experiments in general psychology is a revised and enlarged edition of the author's manual, which appeared in 1929. While the first edition contained only forty-five experiments, the revised edition has eighty-one experiments. Many of them are what we have come to think of as standard psychological experiments; however, several experiments referable to newer fields of research are included. As the author aptly points out, the "performance of the experiments requires no supervision; they are self-directing and can be performed outside the classroom." No special apparatus is called for, and when materials for the various experiments are not included in the manual, they can easily be fabricated by the student. The author has included terse comments on each experiment, which should prove very helpful in the students' observations.

Moreover, as the author indicates, many of the experiments may appear ludicrously simple. Experiments are not invalidated, however, by their simplicity. Some of the great experiments of physics were very simple, but served to motivate profound study.

This manual may be used with any standard textbook of psychology. If, however, the instructor does not use a single text, there is included

on pages xii and xiii "a table of references to topical sources in several widely used texts with which these experiments can be correlated, with provision for additional entries for references in other texts."

It is certainly improbable that any instructor would find it expedient to use all of the experiments listed in the manual. However, they represent a sufficient catholicity so that the instructor may choose those that will illustrate and amplify the topics about which he may wish to organize his course.

PAUL L. WHITELY.

Franklin and Marshall College.

STEVENS, B. *The psychology of physics.* Manchester, England: Sherratt & Hughes, 1939. Pp. xvi+282.

The title of the book and the descriptive notices on the paper cover promise much, but the contents are disappointing. The author believes that psychology reveals some *a priori* necessary intuitive forms of thought and maintains that the conceptual frame of physics can be deduced from those forms. The result is a regression to nineteenth-century ether ideas and other kinds of abandoned "model" theories. The book thus contributes neither to psychology nor to physics.

HERBERT FEIGL.

University of Iowa.

GROOS, H. *Willensfreiheit oder Schicksal?* München: Ernst Reinhardt, 1939. Pp. 277.

This is a rather complete and penetrating analysis of the old free-will puzzle, written in a lively and appealing manner. It is primarily of interest to philosophers, but also those psychologists who are concerned with problems of motivation may find some stimulation here. Although the author seems perfectly clear about the fact that "freedom," in the usual sense of the word, is compatible with determinism, nevertheless, he lapses somehow in concluding that some form of fatalism is inescapable. In any case, the book is one of the best ever written on a problem which, at least according to this reviewer, is so vexatious only because it is engendered by several confusions of meaning.

HERBERT FEIGL.

University of Iowa.

BOOKS RECEIVED

BIERENS DE HAAN, J. A. *Die tierischen Instinkte und ihr Umbau durch Erfahrung: eine Einführung in die allgemeine Tierpsychologie.* Leiden, Holland: E. J. Brill, 1940. Pp. xi + 478.

CANTRIL, H., with the assistance of H. Gaudet & H. Herzog. *The invasion from Mars: a study in the psychology of panic with the complete script of the famous Orson Welles broadcast.* Princeton: Princeton Univ. Press, 1940. Pp. xv + 228.

DOOB, L. W. *The plans of men.* New Haven: Yale Univ. Press, 1940. Pp. xiii + 411.

GRAY, L. H. *Foundations of language.* New York: Macmillan, 1939. Pp. xv + 530.

GUILLAUME, P. *La psychologie animale.* Paris: Armand Colin, 103, Boulevard Saint-Michel, 1940. Pp. 210.

HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940. Pp. xi + 429.

HULL, C. L., HOVLAND, C. I., ROSS, R. T., HALL, M., PERKINS, D. T., & FITCH, F. B. *Mathematico-deductive theory of rote learning: a study in scientific methodology.* New Haven: Yale Univ. Press, 1940. Pp. xii + 329.

JUNG, M. (Ed.) *Modern marriage.* New York: Crofts, 1940. Pp. xiv + 420.

LEWIN, K., LIPPITT, R., & ESCALONA, S. K. *Studies in topological and vector psychology I.* *Univ. Ia Stud. Child Welf.*, Vol. XVI, No. 3. Iowa City: University, 1940. Pp. 307.

MOST, O. J. *Die Determinanten des seelischen Lebens: I. Grenzen der kausalen Betrachtungsweise.* Breslau: Franks Verlag & Druckerei, Otto Borgmeyer, 1939. Pp. 312.

VERNON, P. E. *The measurement of abilities.* London: Univ. London Press, 10 & 11 Warwick Lane, 1940. Pp. xii + 308.

WALTON, A. *The new techniques for supervisors and foremen.* New York: McGraw-Hill, 1940. Pp. vi + 233.

NOTES AND NEWS

DR. EDWARD LEE THORNDIKE, professor of educational psychology and director of the Division of Psychology of the Institute of Educational Research at Teachers College, Columbia University, will retire from active service on July 1.—*Science*.

AT THE meeting of the Society of Experimental Psychologists, held March 26-27 at the University of Pennsylvania, the Howard Crosby Warren Medal was awarded to Ernest R. Hilgard, of Stanford University, for his analysis of the conditioned response and his demonstration of its integration with the verbal and volitional processes in learning and retention. The Warren Medal is awarded annually by the Society for outstanding research in the field of experimental psychology.

DR. GARDNER MURPHY, of Columbia University, has been appointed professor of psychology at the City College, College of the City of New York. The staff of the newly established department of psychology has elected Dr. Murphy to serve as chairman when he assumes his post next September.

As a result of the recent poll held by the New York State Association for Applied Psychology, the following officers have been elected and will assume their duties on July 1: President, Henry E. Garrett; Vice-President, Carney Landis; Treasurer, Arthur L. Benton; Upstate Member of Executive Committee, Warren G. Findley; Metropolitan Member of Executive Committee, W. Douglas Spencer.

AT A meeting of the Washington-Baltimore branch of the American Psychological Association, held January 11, 1940, at the George Washington University, Washington, D. C., the topic "Psychologists in the Government Service" was discussed by the following participants: Dr. Dean R. Brimhall, Civil Aeronautics Authority; Dr. Benjamin Frank, Prison Bureau, Department of Justice; Dr. Carroll Shartle, Federal Security Agency; Dr. Kimball Young, Department of Agriculture; Dr. V. Henmon, Civil Aeronautics Authority; Dr. G. M. Ruch, U. S. Office of Education; Dr. J. P. Shea, U. S. Forest Service.

THE Washington-Baltimore branch of the American Psychological Association met at Howard University on March 14, 1940. The following program was presented:

F. P. WATTS, Howard University: "Comparative Clinical Study of Delinquent and Nondelinquent Negro Boys."

S. M. NEWHALL, Johns Hopkins University: "The Warmth and Coolness of Colors."

F. C. SUMNER, Howard University: "Würzburg *vs.* Psychoanalytic Techniques in the Psychology of Religion."

DR. S. J. BECK, head of the psychology laboratory in the department of neuro-psychiatry at Michael Reese Hospital, will give a course on "The Rorschach Method in Personality Study and Clinical Diagnosis" from June 24 through June 28, 1940. The primary aim of the course will be to demonstrate the test's practical application in investigating the whole personality, with particular reference to its clinical use, and will include teaching the technique of administering the Rorschach Method and scoring and interpreting the responses. Those interested in the course are invited to communicate with the Medical Librarian, Michael Reese Hospital, 2908 Ellis Avenue, Chicago, Illinois.

DR. JERRY W. CARTER, JR., who for the past four years has been senior clinical psychologist at the James Whitcomb Riley Hospital for Children, Indiana University Medical School, Indianapolis, will begin his new duties as consulting psychologist for the Wichita Child Guidance Center on July 1.

A MEETING in memory of the late Margaret Floy Washburn was held at Vassar College on April 14, 1940. The principal address was delivered by President Leonard Carmichael, of Tufts College, representing the American Psychological Association. Dr. Carmichael reviewed Professor Washburn's major contributions to the science of psychology, with special emphasis upon her work in animal psychology. Preceding this address, President Henry Noble MacCracken spoke in appreciation of Miss Washburn's long and distinguished career at Vassar.

The Trustees of Vassar College have announced the establishment of the Margaret Floy Washburn Fund, the income from which will be used to aid promising students, preference being given to students of psychology. Included in this fund are the residual estate of Miss Washburn, bequeathed to Vassar College in her will, and an annuity fund given by her students upon the completion of her twenty-fifth year at that college.

ON June 13 the department of psychology of the University of California at Los Angeles will celebrate the first semester of occupancy of its new building by presenting a program of four scientific papers on topics lying within four important fields of psychology. The speakers are to be: G. M. Stratton, University of California at Berkeley—Social Psychology; Milton Metfessel, University of Southern California—Criminal Psychology; E. R. Hilgard, Stanford University—Experimental Psychology; and R. B. Loucks, University of Washington—Physiological Psychology. In the evening there will be a social gathering of invited guests and members of the Western Psychological Association, which meets in Los Angeles on the following two days.

FROM Boston comes the announcement of a new publication service. The *American Photofile of Psychology* will handle manuscripts for filing with the American Documentation Institute, Washington, D. C., which is equipped to furnish microfilm or photoprint copies at any time through its Bibliofilm Service. The "Photofile" will publish, in reduced facsimile, an author's abstract to be submitted with each original manuscript, guaranteeing circulation to 200 principal psychological centers in the United States as well as to subscribers. Abstracts will appear in a form adapted to card filing, three to a page, thirty to each number of the "Photofile." The service represents an attempt to reduce the costs of publication for material which would not ordinarily find its way into the regular journals. Interested psychologists may communicate with Irving C. Whittemore, Boston University, 685 Commonwealth Avenue, Boston, Massachusetts.

DR. WILLIAM F. BOOK, formerly professor of psychology at Indiana University, died on May 22.

1. *Leucania* *luteola* (Hufnagel) *luteola* (Hufnagel) *luteola* (Hufnagel) *luteola* (Hufnagel)

